

M 2015

Análise de Padrões Sequenciais nos Mercados Financeiros

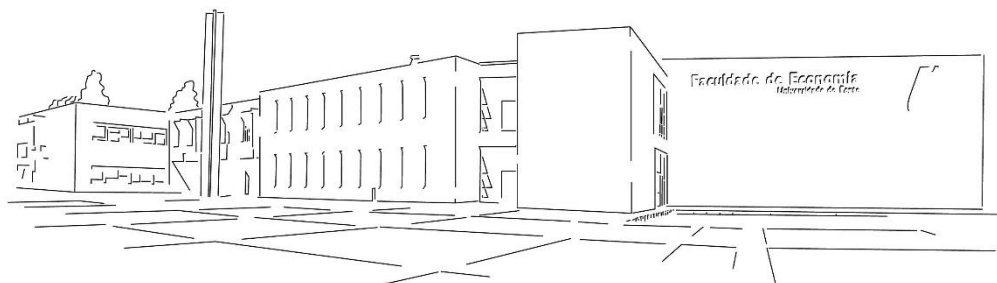
Um estudo sobre a aplicação de *Sequence Pattern Mining* ao movimento de preços das ações

Cátia de Jesus Coimbra Ferreira

DISSERTAÇÃO DE MESTRADO

EM MODELAÇÃO, ANÁLISE DE DADOS E SISTEMAS DE APOIO À DECISÃO

Tecnologia da Informação



Análise de Padrões Sequenciais nos Mercados Financeiros

- Um estudo sobre a aplicação de *Sequence Pattern Mining* ao movimento de preços das ações

por

Cátia de Jesus Coimbra Ferreira

Dissertação de Mestrado em Modelação, Análise de Dados e Sistemas de
Apoio à Decisão

Orientada por:

Prof. Doutor João Manuel Portela da Gama

Prof. Doutor Carlos Abreu Ferreira

Faculdade de Economia

Universidade do Porto

2015

Breve Nota Biográfica

Cátia de Jesus Coimbra Ferreira nasceu em Viseu, em 1992.

Estudou na Universidade de Aveiro onde completou a Licenciatura em Economia em 2013. Ainda nesse ano, por forma a complementar a formação anterior, decidiu ingressar no Mestrado de Modelação, Análise de Dados e Sistemas de Apoio à Decisão, na Faculdade de Economia da Universidade do Porto.

Em 2014, iniciou a atividade profissional numa Instituição Financeira.

Agradecimentos

Quero agradecer a todas as pessoas que, de forma direta ou indireta, contribuíram para a realização desta dissertação.

Em especial, quero agradecer aos meus orientadores, o Professor Doutor João Gama e o Professor Doutor Carlos Ferreira, por todo o apoio, orientação e disponibilidade demonstrado ao longo do último ano.

Este trabalho foi apoiado pela Comissão Europeia no âmbito do projeto MAESTRA (Grant number ICT-2013-612944), a quem fica também o meu agradecimento.

Resumo

Os mercados financeiros são uma das áreas onde mais informação circula diariamente, a qual é produzida pela interação entre os investidores no momento da compra e venda de instrumentos financeiros. Contudo, a análise desta informação torna-se um processo ineficiente quando feito manualmente devido à grande quantidade de dados. O *data mining* é a área de descoberta de conhecimento de forma automática, através de diversas técnicas que visam a exploração de grandes quantidades de dados. O objetivo passa por encontrar, de forma eficiente, padrões/modelos que caracterizem os dados e os transformem em informação relevante para o utilizador. Neste trabalho são analisadas as variações relativas dos preços de fecho de alguns instrumentos financeiros negociados nos mercados financeiros, referentes a um determinado período de tempo, através da técnica de *Sequence Pattern Mining*. A hipótese que se pretende estudar nesta dissertação é que a existência de padrões frequentes implica que alguns comportamentos específicos do mercado financeiro, nomeadamente no preço de fecho dos ativos, ocorram de forma frequente quer ao nível individual (uma empresa) quer ao nível coletivo (grupos de empresas). Por outro lado, pretende-se a identificação de possíveis relações entre os dados que possibilitem conhecimento útil para os utilizadores, nomeadamente através da análise de similaridade entre as empresas/entidades.

Palavras-Chave: Data Mining. Sequence Pattern Mining. Mercados Financeiros.

Abstract

Financial markets are one of the areas where more information circulates daily, which is produced by the interaction between investors at time of buying and selling financial instruments. However, analysis of this information becomes inefficient process when done manually due to the large amount of data. Data mining is the area of knowledge discovery automatically, through various techniques that are aimed at the exploitation of large amounts of data. The goal is to find, efficiently, patterns/models that characterize the data and transform them into relevant information for the user. This work analyzes the relative variations of the closing prices of some financial instruments traded in the financial markets, for a certain period of time, through the technique of Sequence Pattern Mining. The hypothesis to be studied in this thesis is that the existence of frequent patterns implies that some specific behaviors of the financial market, especially in the closing price of assets, occur frequently form at the individual level (a company) or at the collective level (groups of companies). On the other hand, it is intended to identify possible relationships between data that provide useful knowledge for users, in particular through the analysis of similarity between companies/entities.

Keywords: Data Mining. Sequence Pattern Mining. Financial Markets.

Índice

Breve Nota Biográfica	i
Agradecimentos	ii
Resumo	iii
Abstract	iv
1 Introdução	1
1.1 Motivação	2
1.2 Problema a estudar	2
1.3 Estrutura da dissertação	3
2 Trabalho Relacionado	4
2.1 Extração de Conhecimento de Dados	4
2.2 Extração de Conhecimento de Dados Temporais	5
2.3 Extração de Padrões Sequenciais	5
2.3.1 Algoritmos de extração de padrões sequenciais	8
2.3.1.1 Algoritmos baseados na abordagem Apriori	8
2.3.1.2 Crescimento de padrões através de projeções	9
2.3.1.3 Algoritmo SPADE	11
2.3.1.4 Algoritmo cSPADE	14
2.4 Padrões Sequenciais nos mercados financeiros	15
3 Estudo de padrões sequenciais nos mercados financeiros	22
3.1 Definição do problema	22
3.2 Descrição dos dados	22
3.3 Pré-processamento dos dados	24
3.4 Experiências Realizadas	27
3.4.1 Motivação	28
3.4.2 Extração de Padrões Sequenciais	30
3.4.2.1 Extração de sequências frequentes com dados diários	30
3.4.2.2 Extração de sequências frequentes com dados semanais	33
3.4.2.3 Extração de sequências frequentes com restrições	38

	3.4.2.4	Extração de sequências frequentes máximas	43
	3.4.3	Análise de Similaridade	44
	3.4.3.1	Análise de similaridade com dados diários	45
	3.4.3.2	Análise de similaridade com dados semanais	46
	3.4.3.	Estudo individualizado dos padrões sequenciais	48
4		Conclusões e Trabalho Futuro	60
	4.1	Conclusões	60
	4.2	Trabalho Futuro	63
5		Referências Bibliográficas	65
6		Anexos.....	69

Índice de Figuras

2.1	Pseudo-código para a Breadth First Search e para a Depth First Search	13
2.2	Pseudo-código do cSPADE	14
3.1	Tempo de processamento (a) e número de sequências encontradas (b), com dados diários, na discretização mais grossa, com diferentes limites de suporte	32
3.2	Tempo de processamento (a) e número de sequências encontradas (b), com dados diários, na discretização mais fina, com diferentes limites de suporte	33
3.3	Tempo de processamento (a) e número de sequências encontradas (b), com dados semanais, na discretização mais grossa, com diferentes limites de suporte	34
3.4	Tempo de processamento (a) e número de sequências encontradas (b), com dados semanais, na discretização mais fina, com diferentes limites de suporte	35
3.5	Frequência dos itens com dados diários na discretização mais grossa (a) e mais fina (b) e frequência dos itens com dados semanais na discretização mais grossa (c) e mais fina (d)	37
3.6	Tempo de processamento (a) e número de sequências encontradas (b) com restrições adicionais	40
3.7	Frequências dos itens com dados diários na discretização mais grossa (a) e mais fina (b), e com dados semanais na discretização mais grossa (c) e mais fina (d)	41
6.1	Frequências das variações dos preços de fecho na Adidas	92
6.2	Frequências das variações dos preços de fecho na Allianz	93
6.3	Frequências das variações dos preços de fecho na BASF	93
6.4	Frequências das variações dos preços de fecho na Bayer	93
6.5	Frequências das variações dos preços de fecho na BMW	93
6.6	Frequências das variações dos preços de fecho no Commerzbank	94
6.7	Frequências das variações dos preços de fecho no Continental	94
6.8	Frequências das variações dos preços de fecho na Daimler	94
6.9	Frequências das variações dos preços de fecho no Deutsche Bank	94
6.10	Frequências das variações dos preços de fecho na Lufthansa	95
6.11	Frequências das variações dos preços de fecho na Deutsche Telekom	95
6.12	Frequências das variações dos preços de fecho na Fresenius	95
6.13	Frequências das variações dos preços de fecho na Heidelberg Cement	95

6.14	Frequências das variações dos preços de fecho na Linde	96
6.15	Frequências das variações dos preços de fecho na Merck	96
6.16	Frequências das variações dos preços de fecho na Siemens	96
6.17	Frequências das variações dos preços de fecho na Volkswagen	96
6.18	Frequências das variações dos preços de fecho na S&P500	97
6.19	Frequências das variações dos preços de fecho no USA Oil Fund	97

Índice de Tabelas

2.1	Contagem dos itens de uma base de dados	12
2.2	Verificação do suporte dos itens	12
2.3	Id-list para <AB>	13
3.1	Empresas em análise e respectivos setores de atividade	23
3.2	Codificação dos dados	24
3.3	Discretização mais grossa das variações dos preços de fecho	26
3.4	Discretização mais fina das variações dos preços de fecho	26
3.5	Representação dos dados	27
3.6	Sequências frequentes, na discretização mais fina, com dados diários	42
3.7	Sequências frequentes, na discretização mais fina, com dados semanais	42
3.8	Similaridade das entidades com dados diários, na discretização mais fina	46
3.9	Similaridade das entidades com dados semanais, na discretização mais fina	47
6.1	Sequências frequentes máximas com dados diários	70
6.2	Sequência frequentes máximas com dados semanais	88
6.3	Sequências frequentes máximas da Adidas, Allianz e Commerzbank	89
6.4	Sequências frequentes máximas da Continental, Daimler, Deutsche Bank e Deutsche Telekom	89
6.5	Sequências frequentes máximas da Fresenius, Heidelberg, Linde e Merck	90
6.6	Sequências frequentes máximas da Siemens, Volkswagen, S&P500 e USA Oil Fund.....	91
6.7	Sequências frequentes máximas da BMW, da Bayer e da Lufthansa	92

1 Introdução

O desenvolvimento tecnológico das últimas décadas conduziu ao rápido crescimento e circulação da informação. Estes dados, que são armazenados em *data warehouse*, contêm conhecimento importante para as organizações que pode ser usado nas mais diversas estratégias empresariais tais como marketing, investimentos, produção, *cross-selling*, entre outras. Contudo, a extração e exploração manual de grandes quantidades de dados é um processo complexo e, muitas vezes, intratável. O *data mining* tem como objetivo a produção de conhecimento útil, de forma automática, a partir de grandes conjuntos de dados. As diversas técnicas disponíveis nesta área permitem descobrir de forma eficiente padrões/modelos que caracterizam os dados e os transformam em informação útil para o utilizador.

Os mercados financeiros produzem, diariamente, uma enorme quantidade de dados através da interação entre os investidores no momento de compra e venda de instrumentos financeiros. A tomada de decisão dos investidores é limitada pela informação disponível, a qual tem por base métodos tradicionais como a análise técnica e a análise fundamental que, tal como referido em *Marketos et al. (2004)* não possibilitam a exploração de toda a informação disponível. Vários autores defendem o uso de técnicas de data mining nos mercados financeiros como um procedimento eficiente na procura de informação implícita. Porém, de acordo com o estudo em *Fama (1970)*, que defende a teoria/hipótese do mercado eficiente, os preços dos ativos refletem na totalidade toda a informação relevante sobre os mesmos, não admitindo que alguma informação possa não ser conhecida e que permita prever os preços dos ativos. A hipótese que se pretende estudar nesta dissertação é que a existência de padrões sequenciais frequentes implica que alguns comportamentos específicos dos mercados financeiros, nomeadamente no preço de fecho das ações, ocorram de forma frequente quer ao nível coletivo (várias empresas) como ao nível individual (uma empresa). Para tal, será usada a técnica *Sequence Pattern Mining*.

1.1 Motivação

A principal motivação para a realização desta dissertação é descoberta de padrões relevantes presente em dados financeiros temporais. Os padrões sequenciais caracterizam os movimentos dos preços durante o período em análise. Adicionalmente, pretende-se estudar a estabilidade/instabilidade dos preços relativos às entidades e, ainda, a similaridade dos movimentos dos preços. Por outro lado, serão realizadas experiências de forma a analisar o impacto da granularidade temporal dos dados, bem como o efeito da aprendizagem de sequências com restrições nos padrões encontrados.

1.2 Problema a estudar

O objetivo desta dissertação é, a partir de um conjunto de dados referentes a variações dos preços de fecho de dezanove entidades negociadas nos mercados financeiros, extrair informação útil sobre o comportamento/relação desses preços durante um determinado período de tempo. A hipótese a estudar nesta dissertação é que a descoberta de padrões sequenciais frequentes permite identificar que alguns comportamentos específicos nos mercados financeiros, nomeadamente as variações dos preços de fecho dos ativos, ocorram com frequência, quer ao nível coletivo (grupos de empresas) quer ao nível individual (uma empresa).

Nesta dissertação serão analisadas dezanove entidades negociadas nos mercados financeiros, as quais compreendem dezassete ações que compõem o DAX (*Deutscher Aktien Index*) e as outras duas são o índice S&P500 e o USA Oil Fund. Os preços de fecho foram extraídos para o período de Setembro de 2014 a Março de 2015, através do *site* da *Yahoo Finance*. A técnica *Sequence Pattern Mining* será implementada no *software* R.

1.3 Estrutura da Dissertação

A dissertação encontra-se estruturada e organizada da seguinte forma: no capítulo 2 é apresentado o trabalho relacionado onde são abordados alguns conceitos de extração de conhecimento de dados. É feita, ainda, uma introdução à extração de padrões sequenciais, com a apresentação de alguns algoritmos para a realização da mesma, assim como a recolha da literatura relevante. O capítulo 3 inicia-se com a abordagem aos dados utilizados para este estudo, nomeadamente através da descrição e pré-processamento dos mesmos e em seguida será realizado um conjunto de experiências com vista à extração de conhecimento de dados. Por fim, no capítulo 4 são apresentadas as conclusões da dissertação e o trabalho futuro.

2 Trabalho Relacionado

Neste capítulo pretende-se introduzir alguns conceitos relacionados com a extração de conhecimento de dados, nomeadamente no que diz respeito à técnica usada na presente dissertação – extração de padrões sequencias. Serão apresentados alguns dos principais algoritmos bem como a literatura relevante.

2.1 Extração de Conhecimento de Dados

Se por um lado assistimos diariamente a uma grande circulação de informação, por outro lado, o tratamento e utilização da mesma é um procedimento mais complexo. Os dados compreendem diversa informação que pode ser usada para os mais variados fins. Porém, a extração e tratamento da mesma de forma manual é um processo impraticável. A extração de conhecimento de dados (*data mining*) incorpora um conjunto de técnicas e procedimentos que quando aplicados a grandes conjuntos de dados descobrem, de forma automática, informação relevante num curto espaço de tempo.

O processo da descoberta de conhecimento a partir de conjuntos de dados passa desde a forma como esses dados são armazenados, os algoritmos que são aplicados e, ainda, como pode ser obtido, visualizado e interpretado o conhecimento. O processo de descoberta de conhecimento inclui três fases: pré-processamento, a extração dos dados e pós-processamento. A fase de pré-processamento inclui a seleção, limpeza e transformação dos dados, sendo que para tal é necessária a compreensão do domínio da aplicação, ou seja, a partir do objetivo que é proposto será selecionada a informação que possa explicar a hipótese em estudo. A limpeza dos dados é um dos procedimentos mais importantes devido à presença de ruído, inconsistências ou redundância nos mesmos e que impede o bom funcionamento de algumas técnicas de *data mining*. A transformação dos dados consiste na simplificação da sua estrutura de forma a obter um formato mais apropriado para a técnica que será utilizada.

Na segunda etapa é escolhida a técnica/algoritmo de *data mining* mais adequada de acordo com o propósito da análise. Estas técnicas/algoritmos vão aprender, a partir de

um subconjunto de dados (dados de treino), um modelo/hipótese que relacione os dados de entrada com os dados de saída.

Na etapa de pós-processamento é realizada a representação do conhecimento extraído através de modelos de conhecimento, tais como, grupos, gráficos, regras, árvores, entre outros. Além da representação, a interpretação do modelo é um dos aspectos mais importantes para a tomada de decisão e utilização da informação em ações futuras.

A extração de conhecimento de dados permite que seja obtida informação implícita em grandes conjuntos de dados que será, posteriormente, utilizado pelas instituições/utilizadores no processo de tomada de decisão.

2.2 Extração de Conhecimento de Dados Temporais

A extração de conhecimento de dados temporais consiste na obtenção de conhecimento a partir de dados temporalmente ordenados. A principal característica é a presença da dinâmica temporal onde é analisada a forma como a informação está relacionada com as alterações temporais.

Enquanto a extração de conhecimento de dados, referida no ponto 2.1, trata um conjunto de dados desordenados, na extração de conhecimento de dados temporais, o aspecto temporal dos dados é tido em conta no processo de descoberta de conhecimento. O objetivo passa pela descoberta de padrões que caracterizem os dados ao longo de um determinado período de tempo e que permitem a descoberta de comportamentos/procedimentos que ocorrem posteriormente a outros ou, ainda, encontrar comportamentos/procedimentos que se repetem após um determinado intervalo de tempo.

2.3 Extração de Padrões Sequenciais

Um dos tipos de dados mais comuns são as sequências temporais. Estas podem ocorrer em diversos campos tais como na medicina, nas finanças, no comportamento dos consumidores, entre outros. Exemplos de dados sequenciais incluem textos, sequências

ADN, dados da web, compras dos consumidores num supermercado, dados de transações financeiras, entre outros.

Seja $P = \{p_1, p_2, \dots, p_k\}$ um conjunto de itens. Um subconjunto de P é chamado de *itemset*. Uma sequência, s , é definida como $\langle a_1 \rightarrow a_2 \rightarrow \dots \rightarrow a_n \rangle$, onde a_j é um *itemset*, isto é, $a_j \subseteq P$ para $1 \leq j \leq n$. O número de itens numa sequência é chamado de comprimento (*length*) da sequência. Uma sequência de comprimento k é chamada de *k-sequência*.

A extração de sequências frequentes foca-se na descoberta de relações formais entre os dados sequenciais que são designadas de padrões sequenciais. Dado um conjunto de sequências, onde cada sequência é uma lista de transações ordenadas pelo momento da transação (*transaction-time*), e cada transação compreende um conjunto de itens, o problema consiste em encontrar todos os padrões sequenciais que apresentam um determinado suporte mínimo estabelecido pelo utilizador, onde o suporte de um padrão é o número de sequências nos dados que contêm esse padrão (*Srikant et al. 1996*).

O suporte de uma sequência f , num conjunto de sequências F , consiste no número de vezes em que a sequência f ocorre no conjunto de sequências F , denotado como *suporte* (f). Dado um limite mínimo de suporte estabelecido pelo utilizador, p , uma sequência f , é designada de padrão sequencial (sequência frequente) se *suporte* (f) $\geq p$.

$$\text{Suporte}(f) = \frac{\text{Número de ocorrências de } f}{\text{Total de sequências de } F}$$

$$\text{Suporte}(f) \in [0;1]$$

A definição de um suporte mínimo é um processo complexo. Por um lado, um suporte mínimo baixo permite que sejam encontradas um grande número de sequências frequentes o que torna, por vezes, o processo de descoberta de conhecimento intratável, além disso, podem ser descobertas um grande número de sequências que, apesar de verificarem o suporte mínimo, não produzem conhecimento útil. Por outro lado, um suporte mínimo elevado pode conduzir à perda de um número considerável de sequências interessantes do ponto de vista informativo mas que não atingem o suporte mínimo estabelecido. Neste sentido, vários estudos, como em *Antunes et al. (2004)*, *Chakrabarti et al. (1998)* e *Pei et al. (2001b)*, defendem o uso de restrições adicionais para limitar o

número de padrões descobertos e focar o processo de descoberta de conhecimento na informação que possa produzir mais interesse para o utilizador. Este aspeto será discutido na secção 3.4.2.3.

A extração de padrões sequenciais tem como objetivo encontrar subsequências frequentes num conjunto de dados sequenciais. Um padrão sequencial descreve um conjunto de eventos que se sucedem no tempo e que podem ser, posteriormente, utilizados para prever eventos futuros (Agrawal et al. 1995). Um exemplo de um padrão sequencial que pode ocorrer numa base de dados de compras de supermercado é: $\langle \text{água} \rightarrow \text{pão} \rightarrow \text{leite} \rangle$. A partir deste exemplo, é possível prever que, com alguma probabilidade, um cliente que compre água e pão também compra leite. A procura de padrões sequenciais ocorre, assim, sobre uma base de dados transacional onde cada transação identifica um evento.

Vários autores defendem que a extração de padrões frequentes não deve ocorrer sobre todos os padrões frequentes, mas sim sobre os máximos. Tal como definido em Agrawal et al. (1995):

“Given a database D of customer transactions, the problema of mining sequential patterns is to find the maximal sequences among all sequences that have a certain user-specified minimum support. Each maximal sequence represents a sequential pattern.”

Uma sequência $s_a = \langle a_1 \rightarrow a_2 \rightarrow \dots \rightarrow a_n \rangle$ está contida numa outra sequência $s_b = \langle b_1 \rightarrow b_2 \rightarrow \dots \rightarrow b_m \rangle$, se existem inteiros $1 \leq k_1 \leq k_2 \leq \dots \leq k_n \leq m$, tal que $a_1 \subseteq b_{k_1}$, $a_2 \subseteq b_{k_2}$, ..., $a_n \subseteq b_{k_n}$. Denotamos s_a uma subsequência de s_b , e s_b uma supersequência de s_a . Assim, uma sequência é máxima se é frequente e nenhuma das suas supersequências é frequente, ou seja, uma sequência é máxima se não está contida em mais nenhuma sequência frequente.

Note-se que enquanto as regras de associação, introduzidas por Agrawal et al. (1993), incidem sobre os itens de uma transação que ocorrem frequentemente num conjunto, a extração de padrões sequenciais recai sobre a descoberta de itens que ocorrem numa certa ordem.

2.3.1 Algoritmos de extração de padrões sequenciais

Os algoritmos de extração de padrões sequenciais podem ser agrupados em dois tipos: algoritmos baseados na abordagem *Apriori* e algoritmos baseados na metodologia de crescimento de padrões através de projeções. No primeiro grupo encontra-se algoritmos como *AprioriAll* (Agrawal et al. 1995), *GSP* (Srikant et al. 1996), *SPADE* (Zaki, 2001) e *SPAM* (Ayres et al. 2002) e o segundo grupo inclui o *PrefixSpan* (Pei et al. 2001a) ou o *FP-Growth* (Han et al. 2000), entre outros.

2.3.1.1 Algoritmos baseados na abordagem Apriori

Os algoritmos baseados na abordagem *Apriori* têm em comum o uso da estratégia de geração-teste. Esta estratégia consiste em dois passos fundamentais. O primeiro diz respeito à geração de sequências candidatas através da combinação dos *itemsets* frequentes, posteriormente, será realizada a verificação do suporte das mesmas (segundo passo). Este processo é consecutivamente repetido até que não seja possível gerar mais sequências frequentes. Para um melhor entendimento de todo o processo será apresentado o algoritmo GSP que aplica esta estratégia.

- Algoritmo GSP (*Generalised Sequential Patterns*)

Tal como referido anteriormente, em Agrawal et al. (1995) foi introduzido, pela primeira vez, o estudo da extração de padrões sequenciais tendo sido apresentados três algoritmos: *AprioriAll*, *AprioriSome* e *DynamicSome*. O primeiro algoritmo faz parte da chamada família *count-all* e os dois últimos pertencem à família dos *count-some*¹. Existem dois passos essenciais no *AprioriAll*: geração de candidatos e teste. O primeiro passo consiste na geração das sequências que possam ser frequentes sendo, posteriormente, no segundo passo, realizadas passagens na base de dados a fim de verificar o suporte de cada sequência candidata e determinar os padrões sequenciais

¹ Enquanto os algoritmos *count-all* encontram todas as sequências que verifiquem um determinado suporte mínimo, os algoritmos *count-some*, descobrem apenas as sequências máximas.

frequentes. Neste sentido, é fácil testar e contar o suporte dos candidatos, através da verificação direta da base de dados original. Os autores concluíram que o algoritmo *DynamicSome* apresenta um bom desempenho (em termos de tempo de execução e número de sequências encontradas) com suportes mínimos elevados e os algoritmos *AprioriAll* e *AprioriSome* apresentam um desempenho muito semelhante. Contudo, o *AprioriAll* apresenta como limitação o facto de fazer múltiplas passagens sobre os dados e gerar, assim, um grande número de candidatos.

O *GSP* (Srikant et al. 1996) foi proposto um ano mais tarde pelos mesmos autores do *AprioriAll*, sendo, também, um algoritmo baseado na metodologia *Apriori*. Uma das principais diferenças entre o *AprioriAll* e o *GSP* é que este último introduz algumas restrições temporais, como o *maximum gap* e o *minimal gap* que definem a diferença máxima e mínima entre duas quaisquer transações adjacentes numa sequência, e possibilitam, assim, a redução do número de candidatos gerados.

O *GSP*, como algoritmo pertencente à família *Apriori*, inclui os mesmos dois passos no processo de extração de sequências que os algoritmos anteriores: geração de candidatos e teste. Na fase da extração de candidatos, são obtidas as *k*-sequências candidatas baseadas nas (*k*-1)-sequências. Na segunda fase é realizada a contagem do suporte das sequências encontradas, sendo que as que não atingem o suporte mínimo são eliminadas. Os autores concluíram que o *GSP* apresenta um desempenho melhor que o *AprioriAll* pois o número de sequências candidatas é muito menor, devido à incorporação de restrições.

2.3.1.2 Crescimento de padrões através de projeções

A estratégia de crescimento de padrões surge pela necessidade de ultrapassar uma das principais limitações dos algoritmos baseados no *Apriori*: as consecutivas passagens pela base de dados.

Esta metodologia consiste na utilização do método dividir e conquistar que aplica subdivisões nos dados baseando-se na procura de prefixos ou sufixos frequentes que irão fazer com que as sequências “cresçam”. Neste sentido, a exploração das extensões (prefixos e sufixos) das sequências atuais que serão, posteriormente, usadas como “sementes” para o próximo nível, evita a fase de geração e teste de candidatos usada na metodologia Apriori.

Ao contrário da estratégia anterior, na metodologia de crescimento de padrões através de projeções a geração de candidatos é feita em cada subconjunto assim como a contagem do suporte, o que permite a redução significativa do espaço de procura.

As principais vantagens desta metodologia, quando comparada com a anterior, passam pelo facto de reduzir substancialmente o tamanho da base de dados, devido à realização de projeções bem como a capacidade de crescimento da memória principal, tornando assim possível colocar uma parte substancial de dados na memória principal. Contudo, o uso desta metodologia em bases de dados de grande dimensão pode ser um processo pouco eficiente devido ao custo de construção destas projeções. Alguns dos algoritmos que seguem esta metodologia são o *FreeSpan (Frequent Pattern-Project Sequential Pattern Mining)* e o *PrefixSpan (Prefix-Project Sequential Pattern Mining)*.

- Algoritmo PrefixSpan (Prefix-Project Sequential Pattern Mining)

O PrefixSpan (*Pei et al 2001a*) é um dos principais algoritmos baseados no método de crescimento de padrões através de projeções. A metodologia deste algoritmo passa pela exploração dos prefixos das sequências frequentes. Em primeiro lugar é realizada uma passagem pelos dados de forma a encontrar todas as sequências frequentes de comprimento 1. Em seguida, o algoritmo irá dividir o espaço de procura no número de sequências frequentes de comprimento 1 encontradas no passo anterior. Estas sequências frequentes encontradas serão usadas como prefixos para encontrar os restantes padrões sequenciais. Uma das principais características deste algoritmo é que o crescimento dos padrões sequenciais mais longos é feito de forma recursiva a partir das sequências frequentes mais curtas.

2.3.1.3 Algoritmo SPADE

O algoritmo *SPADE* foi proposto por *Zaki (2001)* e pertence à família de algoritmos baseados no *Apriori*. O desenvolvimento deste algoritmo passou pela tentativa de ultrapassar um dos principais problemas apresentados pelos algoritmos da família *Apriori*, nomeadamente no que diz respeito às consecutivas passagens pelas bases de dados. O *SPADE*, ao contrário dos restantes algoritmos *Apriori*, usa propriedades combinatórias para decompor o problema original em subproblemas menores, que podem ser independentemente resolvidos na memória principal usando técnicas de pesquisa *lattice*, e usando operações de agregação. Através deste processo, o algoritmo executa apenas três passagens pelos dados. O primeiro passo consiste na passagem pela base de dados a fim de encontrar todas as 1-sequências frequentes. Em seguida, é realizada uma segunda verificação dos dados para descobrir as 2-sequências frequentes. Com as sequências de comprimento 2 anteriormente encontradas, é construída a estrutura *lattice* (rede). Durante a terceira passagem pela base de dados, todas as sequências grandes são enumeradas através do uso de junções temporais.

Existem dois métodos principais para enumerar sequências frequentes de uma classe: a procura em largura (*BFS - breadth first search*) e a procura em profundidade (*DFS - depth first search*). Na *breadth first search*, em primeiro lugar são geradas todas as sequências de tamanho 1 (1-sequência), por exemplo $\langle A \rangle$, $\langle B \rangle$, $\langle C \rangle$, depois a partir das 1-sequências candidatas são construídas as 2-sequências $\langle AB \rangle$, $\langle A \rightarrow C \rangle$, e assim sucessivamente. De forma contrária, na *depth first search* primeiro é gerada e contada a sequência $\langle A \rangle$, depois, se for frequente, é gerada e contada a sequência $\langle AB \rangle$, e assim sucessivamente. O método BFS apesar de necessitar de mais memória principal que o DFS para guardar todas as 2-sequências consecutivas, dispõe de muito mais informação para podar as sequências de comprimento k que o segundo método. Contudo, o segundo método apenas necessita de guardar a última sequência de comprimento 2 e as novas sequências geradas de comprimento k (*Zaki, 2001*). A enumeração das sequências frequentes pode ser realizada através da agregação das *id-lists*² em três formas, tal como indicado em *Zaki, (2001)*, assumindo que X e Y são itens e S é uma sequência:

² As *id-lists* representam listas para cada item/sequência que contêm o SID (identificador da transação) e o EID (identificador no momento no qual ocorre o evento).

- (a) Itemset e Itemset: agregar XS e YS resulta na sequência XYS;
- (b) Itemset e Sequência: agregar XS com $Y \rightarrow S$ resulta na sequência $Y \rightarrow XS$;
- (c) Sequência e Sequência: agregar $X \rightarrow S$ com $Y \rightarrow S$, resulta em três casos: um novo itemset $XY \rightarrow S$, e duas novas sequências $X \rightarrow Y \rightarrow S$ e $Y \rightarrow X \rightarrow S$.

Na figura 2.1 é apresentado o pseudo-código para a *Breadth First Search* e para a *Depth First Search*. O *input* do procedimento corresponde a um conjunto de “átomos”³ de uma *sub-lattice* S. As sequências frequentes são geradas através de junções das *id-lists* de todos os pares de átomos e da verificação da cardinalidade da *id-list* resultante através do suporte mínimo. Note-se que antes da junção das *id-list*, deve ser inserida uma etapa de poda para garantir que todas as subsequências das sequências resultantes são frequentes. Posto isto, o próximo passo consiste na junção da *id-list*. As sequências frequentes que são encontradas num determinado nível são, posteriormente, usadas na formação dos átomos de classes no próximo nível, sendo o processo repetido de forma recursiva. Por exemplo, vamos considerar a seguinte tabela com a contagem dos itens numa determinada base de dados:

Itens	Nº de Ocorrências
A	4
B	4
C	1
D	2
E	4

Tabela 2.1: Contagem dos itens de uma base de dados

Considerando um suporte mínimo de 2, o próximo passo consiste em verificar os itens que apresentam o suporte mínimo, tal como na tabela seguinte:

Itens	Número de ocorrências
A	4
B	4
D	2
E	4

Tabela 2.2: Verificação do suporte dos itens

³ As sequências que são encontradas no nível atual formam os átomos das classes do próximo nível.

O próximo passo consiste em fazer agregações temporais dos itens para encontrar seqüências frequentes. Por exemplo, a tabela seguinte é uma id-list dos itens $\langle AB \rangle$.

$\langle AB \rangle$			
SID	EID	A	B
1	15		15
1	20		20
2	15		15
3	10		10

Tabela 2.3: Id-list para $\langle AB \rangle$

Neste caso, $\langle AB \rangle$ ocorre quatro vezes no mesmo momento (EID), contudo, a seqüência ocorre duas vezes no mesmo SID (identificador da seqüência) e, portanto, um deles é ignorado. Assim, a frequência de $\langle AB \rangle$ é de 3, estando, portanto, acima do suporte mínimo. Este processo é repetido para $\langle AD \rangle$, $\langle AE \rangle$, $\langle A \rightarrow B \rangle$, $\langle A \rightarrow D \rangle$, e assim sucessivamente.

```

Enumerate-Frequent-Seq (S):
  for all atoms  $A_i \in S$  do
     $T_i = \emptyset$ ;
    for all atoms  $A_j \in S$ , with  $j \geq i$  do
       $R = A_i \vee A_j$ ;
      if (Prune(R) == FALSE) then
         $\mathcal{L}(R) = \mathcal{L}(A_i) \cap \mathcal{L}(A_j)$ ;
        if  $\sigma(R) \geq \text{min\_sup}$  then
           $T_i = T_i \cup \{R\}$ ;  $\mathcal{F}_{[R]} = \mathcal{F}_{[R]} \cup \{R\}$ ;
      end
    if (Depth-First-Search) then Enumerate-Frequent-Seq ( $T_i$ );
  end
if (Breadth-First-Search) then
  for all  $T_i \neq \emptyset$  do Enumerate-Frequent-Seq ( $T_i$ );

```

Figura 2.1: Pseudo-código para a Breadth First Search e para a Depth First Search (retirando de Zaki (2001))

Zaki (2001) concluiu que o *SPADE* apresenta um desempenho melhor que os algoritmos anteriores da família *Apriori*. Em Slimani et al. (2013) os autores mostram que o *SPADE* apresenta um desempenho melhor que o *GSP* devido ao facto do primeiro usar uma metodologia de contagem de suporte mais eficiente baseado numa estrutura *id-list*.

2.3.1.4 Algoritmo cSPADE

O algoritmo *cSPADE*, introduzido por *Zaki (2000)*, apresenta um procedimento em tudo semelhante ao *SPADE*, sendo que a principal diferença ocorre ao nível da incorporação de restrições adicionais que este último não inclui. Estas restrições que o *cSPADE* permite definir passam pela (a) limitação do comprimento das sequências; (b) implementação de limites temporais máximos e/ou mínimos para a ocorrência de elementos consecutivos numa sequência e (c) definição de janelas temporais na descoberta de sequências, o que determina o intervalo máximo de ocorrência de uma sequência frequente.

Face ao exposto, o algoritmo *cSPADE* foi o escolhido para o estudo dos padrões sequenciais nesta dissertação. O pseudo-código para o mesmo pode ser verificado na figura 2.2.

```

cSPADE ( $\text{min\_sup}$ ):
     $\mathcal{P} = \{\text{parente classes } \mathcal{P}_i\};$ 
    for each parente class  $\mathcal{P}_i \in \mathcal{P}$  do Enumerate-Frequent ( $\mathcal{P}_i$ );
Enumerate-Frequent ( $S$ ):
    for all sequences  $A_i \in S$  do
        if ( $\text{maxgap}$ ) // join with =  $\mathcal{F}_2$ 
             $p = \text{Prefix-Item}(A_i);$ 
             $N = \{\text{all 2-sequences } A_j \text{ in class } [p]\}$ 
        else // self-join
             $N = \{\text{all sequences } A_j \in S, \text{ with } j \geq i\}$ 
        for all sequences  $\alpha \in N$  do
            if ( $\text{length}(R) \leq \text{max}_l$  and  $\text{width}(R) \leq \text{max}_w$ 
                and  $\text{accuracy}(R) \neq 100\%$ )
                 $\mathcal{L}(R) = \text{Constrained-Temporal-Join}(\mathcal{L}(A_i),$ 
                     $\mathcal{L}(\alpha), \text{min\_gap}, \text{max\_gap}, \text{window});$ 
                if ( $\sigma(R, c_i) \geq \text{min\_sup}(c_i)$ ) then
                     $T = T \cup \{R\};$  print  $R;$ 
            Enumerate-Frequent ( $T$ );
    Delete  $S$ ;

```

Figura 2.2: Pseudo-código do cSPADE (retirado de Zaki (2000))

2.4 Padrões Sequenciais nos Mercados Financeiros

Apesar da descoberta de regras sequenciais ter surgido para as transações de produtos nas compras de supermercado, a pesquisa foi rapidamente alargada a outras áreas, como a financeira. Entretanto, diversos trabalhos foram desenvolvidos na área de extração de padrões sequenciais.

Vários autores têm aplicado a extração de padrões sequenciais nos mercados financeiros. Os métodos e procedimentos efetuados são muito diversos, contudo, convergem no objetivo de fornecer uma ferramenta capaz de apoiar os investidores no seu processo de tomada de decisão, como é o caso do estudo desenvolvido por *Marketos et al.* (2004) que consiste no desenvolvimento de uma solução para a gestão de carteiras de investimento. Neste trabalho são usadas diversas variáveis que influenciam o preço das ações como o preço de fecho, de abertura, o máximo, o mínimo e o volume transacionado, a fim de retirar o máximo de informação possível. Este sistema consiste num conjunto de técnicas que fazem a extração de informação online e, em seguida, armazena essa mesma informação localmente. Estas ferramentas são otimizadas de modo a fazer várias verificações nos documentos para evitar repetições e diminuir o espaço de procura. Numa primeira fase é feita uma monitorização em que, através do acesso a notícias e/ou através da definição de alertas com base na ocorrência de determinados eventos, o utilizador consegue construir/alterar a sua carteira de investimentos. Na segunda fase é feita a previsão dos preços das ações, através de um algoritmo de extração de sequências frequentes com base na flutuação do preço das ações e, assim, o sistema irá fazer recomendações ao utilizador para comprar ou venda dos ativos. Uma outra funcionalidade disponibilizada por este sistema é a análise de cenários (também conhecida pela análise *What-If*), ou seja, através dos padrões sequenciais encontrados é possível perceber o que pode acontecer quando determinados eventos ocorrem, por

exemplo, “O que acontece quando o Banco Central Europeu decide subir a taxa de juro de referência?” ou “O que acontece aos mercados financeiros quando o preço do petróleo sobe drasticamente?”. Tal como sugerido por *Das et al. (1998)*, a previsão é feita ao nível da flutuação dos preços, que ocorre num determinado intervalo de tempo, ao invés de produzir uma previsão pontual. As regras descobertas são apresentadas do seguinte modo: $A \rightarrow^T B$ (Se A ocorre, então B irá ocorrer dentro de um intervalo de tempo T).

Noutros trabalhos é combinada a análise técnica, através da análise gráfica, com a extração de padrões sequenciais, como no trabalho desenvolvido por *Parracho et al. (2010)*. O objetivo passa por prever a evolução futura de determinada ação. O processo é dividido em dois passos, sendo que na primeira parte é feita a identificação dos padrões com base no preço histórico das ações. Na segunda fase, são utilizados algoritmos genéticos e de computação evolutiva a fim de otimizar os parâmetros do modelo. Para a descoberta de padrões são usados quatro parâmetros de modo a reduzir o número de padrões gerados. Estes parâmetros são: o tamanho da janela deslizante, a taxa de eliminação de ruído, o *Fit Buy* e o *Fit Sell*. Numa série temporal com milhares registos/observações, a janela deslizante irá limitar a análise a um determinado intervalo temporal que corresponde ao número de dias de negociação e, em seguida, será ajustada ao intervalo temporal posterior. Em cada janela, o ruído é eliminado através de uma determinada taxa definida pelo utilizador, sendo que esta taxa está compreendida entre 0% e 50%. O *Fit Buy* indica a presença de um padrão com tendência de preço ascendente, enquanto o *Fit Sell* apresenta um ponto de inflexão, ou seja, um ponto onde já não existe tendência ascendente no preço das ações. Estes quatro parâmetros ao serem definidos pelo utilizador limitam o número de padrões gerados e, tornam, assim, o algoritmo mais eficiente. Para a otimização destes parâmetros é utilizada a técnica *genetic Optimization Kernel I*. O processo passa pela identificação dos padrões de tendência de preço ascendente e, para estes casos, a ferramenta vai sugerir ao investidor que compre a ação, sendo que no momento em que é detetado um ponto de inflexão o investidor vai, de imediato, vender a sua posição e só volta ao mercado quando existir outra tendência de preço ascendente. Esta ferramenta foi testada em três índices de mercado: *S&P500*, *Dow Jones Industrial Average* e *NYSE Composite Index*. Das experiências realizadas ficou demonstrado que a ferramenta proposta, quando aplicada a índices de mercado, produz

melhores resultados que a estratégia *Buy and Hold*⁴. Contudo, quando é aplicada em ações, os resultados não são tão satisfatórios. A justificar esta situação está o facto de os índices de mercado representarem valores médios de um determinado grupo de ações, além disso os índices utilizados nas experiências são índices considerados de baixa volatilidade, o que torna o processo mais eficiente. Para ultrapassar este obstáculo, os autores decidiram aplicar uma nova função Genetic Optimization Kernel II às ações. A alteração dos parâmetros utilizados anteriormente foi decisivo para o sucesso da ferramenta. Conclui-se, assim, que o algoritmo é capaz de detetar com boa precisão a presença de tendência de aumento do preço e, consequentemente, sugerir ao investidor a compra quando tal deve acontecer, assim como, sugere a venda quando deteta a inversão da tendência.

A previsão nos mercados financeiros é complexa e nem todos os métodos são capazes de fazer previsão de modo eficiente devido à influência de fatores externos, como notícias. Para resolver este problema, *Leleu et al.* (2002) sugerem uma técnica que tem como objetivo sinalizar eventos, representados por sequências, que ocorrem nos mercados financeiros. A diferença em relação a outros métodos diz respeito à possibilidade de relacionar a ocorrência de determinados eventos, que aqui estão associados a notícias, a determinado impacto na tendência do preço das ações (aumento, diminuição ou estabilidade). Ao contrário de outros trabalhos, este apresenta a vantagem de poder integrar informações quantitativas e qualitativas. As notícias são fornecidas por agências especializadas em finanças, como a Reuters ou a Bloomberg, e apresentam-se sob a forma de série temporal. Por outro lado, os dados sobre os preços das ações encontram-se, também, representados numa série temporal que foi previamente discretizada em três categorias: “tendência de aumento”, “tendência de diminuição” e “tendência de estabilização”. As duas séries temporais (notícias e tendências) são agrupadas numa só de modo a associar as notícias às tendências de movimentação dos preços. A interpretação da mesma seria, por exemplo, “ocorre uma diminuição significativa no preço das ações da empresa X, quando foi anunciada uma fusão da empresa Y e Z”. Este trabalho foca-se na procura de padrões característicos, e não apenas padrões frequentes. A diferença reside no facto de que um padrão característico não inclui

⁴ Compra de ativos de empresas que apresentam bons resultados económico-financeiros.

apenas padrões frequentes mas, também, padrões raros que podem estar associados a algum evento específico. Tal como noutros trabalhos, o uso de restrições neste tipo de análise é fundamental para reduzir o número de padrões que são descobertos e, assim, para tornar o processo mais eficiente. Contudo, o uso excessivo de restrições pode limitar o espaço de procura e impedir que determinados padrões, interessantes mas que não frequentes, sejam descobertos. Neste sentido, a quantidade de restrições usadas deve ser um aspeto muito ponderado. Os autores optaram pela aplicação do algoritmo *cSPADE* pois permite a integração de diversas restrições como, por exemplo, a inclusão/exclusão de determinados itens. A metodologia passa pela utilização de uma base de dados real com as tendências dos preços (aumento, diminuição, estabilidade) já identificadas, em seguida, o algoritmo é aplicado a cada uma destas classes e são realizadas diversas experiências com diferentes limites de frequência (suporte). O tempo de execução do algoritmo é razoável. Contudo, os autores concluíram que os padrões que têm uma duração muito longa não são os mais relevantes, assim sendo, é necessário fazer o ajustamento da janela, por outro lado, a utilização de um limite de frequência elevado não produz melhores padrões característicos.

Num outro estudo sobre a previsão da tendência do preço das ações é aplicado o algoritmo *AprioriAll* e a técnica *K-means*, como em *Wu et al.* (2014). O procedimento passa pela extração de padrões frequentes e, posteriormente, a previsão do preço das ações. Numa primeira parte é feita a divisão da série temporal, que contém os preços das ações, em gráficos através de janelas deslizantes. Em seguida, estes gráficos são agrupados através do algoritmo *K-means* de modo a formarem padrões gráficos e, posteriormente, a aplicação do algoritmo *AprioriAll* fornece os padrões frequentes. A aplicação deste estudo a uma base de dados real foi feita usando os dados do *Taiwan Stock Exchange Corp.* Nas experiências realizadas foram testados vários parâmetros, como o tamanho da janela deslizante, o suporte e a confiança. Tal como noutros trabalhos, os autores concluíram que para um tamanho de janela deslizante grande os padrões sequenciais descobertos iriam ser muito longos o que poderia dificultar a aplicação do algoritmo *K-means*, além disso a precisão tende a baixar à medida que o tamanho aumenta. Assim, a combinação escolhida foi um tamanho de janela deslizante 20, o suporte mínimo de 13 e a confiança de 10%. Concluiu-se que o método é capaz de fazer boas previsões da tendência das ações no mercado financeiro.

Approximation and Prediction of Stock Time-Series Data (APST) é o método desenvolvido por Vishwanath *et al.* (2013) e tem como objetivo a previsão da tendência do preço das ações. À semelhança de outros trabalhos, também neste é usada a janela deslizante (*sliding windows*) para dividir a série temporal, que contém os preços de fecho históricos de várias empresas. Na primeira parte é usada a técnica *Multilever Segment Mean (MSM)* como pré-processamento dos dados de modo a gerar as sequências que serão, posteriormente, usadas na fase de previsão. Para o processo de previsão é usado o algoritmo *Nearest Neighbors* que irá encontrar os *k* elementos dos vizinhos mais próximos sendo, posteriormente, calculadas as médias dos *k*-elementos. A sequência dos preços de ações prevista é, assim, estimada a partir da média das sequências encontradas no passo anterior. Este modelo é, posteriormente, aplicado a dois conjuntos de dados reais: *Taliwan Stock Exchange Index* e *Bombay Stock Exchange Index* em diferentes empresas. Para a avaliação dos resultados foram usadas duas medidas: *Mean Error Relative (MER)* e o *Mean Absolute Error (MAE)*. Ao comparar este modelo (*APST*) com o *LBF* apresentado por Martínez-Álvarez *et al.* (2008), conclui-se que o método *APST* produz melhores resultados.

Leigh *et al.* (2002) usam gráficos heurísticos (*charting heuristics*) para prever os preços das ações. A análise técnica, tal como referido anteriormente, usa gráficos que contêm o preço e o volume histórico das ações. Neste trabalho são usados este tipo de gráficos aplicados a uma heurística para fazer correspondência a regras de negociação da forma: “Se é identificado o padrão gráfico *X* em *N* dias de negociação anteriores, então comprar; e vender nos *Y*-ésimos dias de negociação seguintes”. Ao contrário de outros trabalhos em que são identificados diversos tipos de padrões, neste é identificado apenas um tipo: *bull flag*⁵. É usada uma matriz 10x10 para identificar a ocorrência das duas variações do padrão gráfico *bull flag*, usando uma janela deslizante de 120 dias. A primeira coluna da matriz representa os 10% primeiros dias de negociação, ou seja, os primeiros 12 dias dos 120, a segunda coluna tem os 10% dos seguintes dias e assim sucessivamente. Os dados correspondem ao preço de fecho do NYSE de 6 de Agosto de 1980 a 8 de Junho de 1999, ou seja, o correspondente a 4748 dias de negociação.

⁵ O *bull flag* designa um padrão gráfico que se assemelha a uma bandeira com um mastro de cada lado, que mostra uma consolidação dentro da tendência e resulta de flutuações de preços dentro de uma faixa, sendo que este tipo de padrão é precedido e seguido por fortes aumentos ou quedas.

Concluiu-se que a ferramenta seleciona eficazmente os dias de compra e melhora a rentabilidade da carteira para esses dias identificados.

Outros trabalhos usam a descoberta de padrões para sinalizar situações em que o investidor deve comprar ações no mercado financeiros, como é o caso do estudo apresentado por Wang *et al.* (2009). Os autores focam a análise na descoberta de dois padrões típicos nos mercados financeiros: *rounding top* e *saucer*. A escolha destes dois padrões deve-se ao facto de geralmente após a ocorrência de um deles existir movimentos substâncias nos preços o que pode levar o investidor a registar ganhos significativos. O *saucer* desenvolve-se quando o mercado atinge preços muito baixos, sendo que a sua forma assemelha-se à letra “U”. O preço cai até atingir um mínimo, seguidamente a situação inverte-se e os preços crescem substancialmente. O *rounding top* apresenta um comportamento oposto ao *saucer*. As variações ocorridas nos preços são mapeadas para uma matriz e através desta sempre que seja identificado um destes padrões, o sistema produz regras de negociação, emite alerta de compra e sinaliza situações de saída de mercado quando é identificado um ponto de reversão da tendência. As regras são definidas a partir de limites, ou seja, sempre que o peso na matriz seja superior a um determinado limite então o investidor deve comprar e vender q dias depois. O modelo foi, posteriormente, aplicado ao índice de mercado *NASDAQ* e os autores concluíram que o mesmo apresenta um poder explicativo relevante quando comparado com outras tecnologias de negociação de ações nos Estados Unidos.

No estudo desenvolvido por Zhang *et al.* (2010) o objetivo passa por construir um algoritmo baseado no coeficiente de Spearman, num conjunto de regras e na janela deslizante, de modo a fazer uma correspondência eficiente entre padrões e as séries temporais. O uso do coeficiente de Spearman tem como finalidade a classificação dos padrões preferenciais, enquanto as janelas deslizantes permitem a redução do tempo de processamento e o conjunto de regras fornece uma consulta de padrões mais eficiente e eficaz. A identificação dos padrões é feita através da combinação da janela deslizante e da técnica dos Pontos Visualmente Importantes (PIP). Esta técnica permite identificar pontos característicos com base nas flutuações dos preços e define os pontos mais distantes (com maiores flutuações) como PIP's. O algoritmo funciona do seguinte modo: os dois primeiros pontos da sequência temporal são definidos como o primeiro e último ponto da sequência *input* P. O próximo ponto PIP será um ponto em P que tenha uma

distância máxima aos primeiros dois pontos PIP's. Com a ajuda da janela deslizante definida para um determinado comprimento, o algoritmo irá percorrer toda a série temporal de modo a descobrir padrões. A utilização do coeficiente de *Spearman* para classificar os padrões descobertos apresenta algumas vantagens como o facto de este critério poder ser utilizado com diferentes tipos de padrões, com intervalos temporais diferentes e não é sensível a *outliers*. Por outro lado, apresenta a desvantagem de por vezes perder alguma informação quando os dados são convertidos em classificações. Para a avaliação do desempenho do método proposto são usados outros dois métodos: a distância euclidiana e o método baseado no declive. Ao comparar os resultados para a previsão concluiu-se que o método proposto neste trabalho apresenta um bom desempenho.

3 Estudo de padrões sequenciais nos mercados financeiros

Neste capítulo será realizada a extração de conhecimento a partir do conjunto dos dados financeiros. Inicia-se com a preparação dos dados e segue-se com diversas experiências com vista a extrair informação relevante para o utilizador.

3.1 Definição do Problema

O objetivo desta dissertação é, a partir de um conjunto de dados referentes a variações dos preços de fecho de dezanove entidades negociadas nos mercados financeiros, extrair informação útil sobre o comportamento/relação desses preços durante um determinado período de tempo. A hipótese a estudar nesta dissertação é que a descoberta de padrões sequenciais frequentes permite identificar que alguns comportamentos específicos nos mercados financeiros, nomeadamente as variações dos preços de fecho das ações, ocorrem com frequência, quer ao nível coletivo (grupos de empresas) quer ao nível individual (uma empresa). Será realizada a discretização dos preços por intervalos de variação de forma a serem, posteriormente, aplicados ao cSPADE. O objetivo consiste na extração de padrões da forma, por exemplo, $\langle A \rightarrow E \rightarrow D \rangle$, onde, com determinada frequência, ocorre uma variação do preço de fecho superior a 2% (categoria “A”), seguindo-se uma variação entre -1% e 1% (categoria “E”) e que, posteriormente, conduz a uma variação do preço de fecho inferior a -2% (categoria “D”). Esta tarefa será realizada, numa primeira fase, no conjunto das dezanove entidades em estudo e, posteriormente, de forma individual, em cada entidade.

3.2 Descrição dos dados

O primeiro passo para a extração e análise de padrões sequenciais nos mercados financeiros consiste na seleção dos dados e construção da respetiva base de dados. Para tal, procedeu-se à recolha de dados relativos ao preço de fecho de dezassete ações negociadas na Bolsa de Valores de Frankfurt, bem como o preço de fecho do índice de

mercado *Standard&Poor 500* e o preço de fecho do fundo de investimentos *USA Oil Fund*, através do site da *Yahoo Finance*. O período em análise é de sete meses, compreendidos entre Setembro de 2014 e Março de 2015.

O DAX (*Deutscher Aktien Index*) é composto pelas 30 maiores empresas alemãs negociadas na bolsa de valores de Frankfurt. Sendo a Alemanha a maior economia da Europa e tendo em conta que as 30 empresas que compõem o índice representam 75% do valor total da bolsa de valores de Frankfurt, o DAX é um dos principais índices de referência europeus. Importa ainda referir que grande parte das empresas negociadas no DAX também são negociadas nos EUA. Neste sentido, a variação dos preços das mesmas pode servir como indicador das tendências nos mercados americanos. As dezassete empresas pertencentes a este índice e escolhidas para a análise constam na tabela 3.1.

Empresa	Setor de atividade
Adidas	Bens de uso pessoal
Allianz	Seguros de não vida
BASF	Produtos químicos
Bayer	Produtos químicos
BMW	Automóveis e componentes
Commerzbank	Banca/setor financeiro
Continental	Automóveis e componentes
Daimler	Automóveis e componentes
Deutsche Bank	Banca/setor financeiro
Deutsche Lufthansa	Viagens e Lazer
Deutsche Telekom	Telecomunicações
Fresenius SE&CO	Serviços e equipamentos de cuidados de saúde
HeidelbergCement	Construção e materiais
Linde	Produtos químicos
Merck	Farmacêuticas e biotecnologia
Siemens	Produtos industriais gerais
Volkswagen	Automóveis e componentes

Tabela 3.1: Empresas em análise e respetivos setores de atividade

O índice *S&P500* (da *Standard and Poor*) é composto por 500 empresas e é o segundo índice mais citado, depois do *Dow Jones Industrial Average*. O *S&P500* representa cerca de 70% do valor total dos mercados de ações dos EUA. Devido ao seu tamanho e diversidade o *S&P500* é referido, e termos gerais, como o melhor indicador quando comparado com o *Dow Jones Industrial Average*.

O United States Oil Fund é um fundo de investimento negociado no NYSE Arca que reflete a variação dos preços dos contratos futuros do petróleo no mercado internacional. Estes contratos não agregam o preço da matéria-prima em si, mas sim o preço do contrato futuro que apresenta uma determinada maturidade. É um fundo cotado que pretende acompanhar a evolução do petróleo.

O objetivo de incluir, na mesma análise, os preços de fecho do índice S&P500, do fundo de investimentos e de um grupo de ações negociadas no DAX passa por tentar descobrir possíveis relações na variação dos preços entre ativos financeiros europeus e norte-americanos.

3.3 Pré-processamento dos dados

Após a recolha dos dados relativos ao preço de fecho das dezassete empresas negociadas no DAX, do índice de mercado S&P500 e do USA Oil Fund para o período de Setembro de 2014 a Março de 2015, o próximo passo consiste na preparação e tratamento destes a fim de serem, posteriormente, aplicados ao algoritmo de extração de padrões sequenciais.

Em primeiro lugar, será realizada a codificação dos dados através da numeração da designação das entidades de forma a facilitar a compreensão e aplicação das técnicas de *data mining*. A codificação será efetuada conforme consta na tabela 3.2.

Código	Empresa/Entidade	Código	Empresa/Entidade
1	Adidas	11	Deutsche Telekom
2	Allianz	12	Fresenius SE&CO
3	BASF	13	HeidelbergCement
4	Bayer	14	Linde
5	BMW	15	Merck
6	Commerzbank	16	Siemens
7	Continental	17	Volkswagen
8	Daimler	18	S&P500
9	Deutsche Bank	19	USA OIL Fund
10	Deutsche Lufthansa		

Tabela 3.2: Codificação dos dados

Em seguida, visto que os preços de fecho se encontram definidos em valores absolutos, que no caso das empresas do DAX são euros e no caso do S&P500 e do USA Oil Fund são dólares, será efetuada a transformação desses valores em variações relativas, usando a seguinte formula:

$$\text{Variação relativa (t)} = \frac{P_t - P_{t-1}}{P_{t-1}}$$

em que P_t define o preço de fecho do instrumento financeiro no instante t e P_{t-1} define do preço de fecho do instrumento financeiro no instante $t-1$.

Em suma, os dados usados para esta análise correspondem as variações relativas do preço de fecho de dezassete empresas negociadas no DAX, do preço de fecho do S&P500 e do USA Oil Fund, durante o período compreendido entre Setembro de 2014 e Março de 2015.

Um dos objetivos do estudo em causa consiste na análise da granularidade temporal dos dados. Para tal pretende-se analisar o impacto na extração de padrões sequenciais do uso de variações diárias dos preços e de variações semanais dos preços. Os dados diários consistem nas variações relativas entre o preço de fecho no dia t e o preço de fecho no dia $t-1$. Com dados semanais, é calculada a variação relativa entre o preço de fecho na semana t e o preço de fecho na semana $t-1$. Adicionalmente, além da realização de experiências com variações diárias e semanais, os dados serão discretizados em intervalos de variação. Esta discretização será realizada em dois tipos.

No primeiro tipo são estabelecidas três categorias de variação (tabela 3.3). Para variações do preço de fecho superiores a 1% é atribuída a categoria “A”, com variações inferiores a -1% é indicada a categoria “D” e, por fim, as variações do preço de fecho entre -1% e 1% são categorizadas pela letra “E”. Este primeiro tipo será designado, nas experiências futuras, por discretização mais grossa.

No segundo tipo pretende-se estudar as variações superiores a 1% e inferiores a -1% (tabela 3.4). Assim, serão adicionadas mais duas categorias por forma a identificar as variações que ocorrem entre 1% e 2%, para as quais foram atribuídas a letra “B”, e

as variações entre -2% e -1%, identificadas com “C”. A letra “E” mantém a identificação às variações ocorridas entre -1% e 1%, sendo que às variações mais expressivas são atribuídas as letras “A” e “D” (variações superiores a 2% e inferiores a -2%, respetivamente). Este tipo será designado, nas experiências futuras, por discretização mais fina.

A utilização de dois tipos de discretização tem como objetivo o estudo do impacto da granularidade de discretização nos resultados obtidos.

Categoria	Variação relativa do preço de fecho
A	>1%
E	[-1%; 1%]
D	<-1%

Tabela 3.3: Discretização mais grossa das variações dos preços de fecho

Categoria	Variação relativa do preço de fecho
A	$\geq 2\%$
B	[1%; 2%[
E	[-1%; 1%[
C	[-2%; -1%[
D	< -2%

Tabela 3.4: Discretização mais fina das variações dos preços de fecho

O estudo da granularidade temporal tem como objetivo avaliar o impacto do uso de variações diárias ou semanais na extração de conhecimento a partir dos padrões sequenciais. Pretende-se responder a questões como: será que o uso das variações semanais dos preços produz mais conhecimento que o uso das variações diárias? Ou não existe qualquer diferença na informação produzida? Ou ainda, qual o impacto da alteração do tipo de discretização na tarefa de extração de padrões sequenciais?

Após a recolha e codificação dos dados o próximo passo consiste na representação dos mesmos a fim de serem, posteriormente, aplicados ao algoritmo. Os dados são representados através de quatro componentes: *sequenceID*, *eventID*, *Size* e *Item*. A *sequenceID* identifica a empresa/entidade onde ocorre a variação do preço de fecho. Esta componente foi codificada de 1 a 19 e corresponde às 19 entidades incluídas nesta análise (ver tabela 3.2). O *eventID* indica o momento em que foi registada a variação no preço

para cada entidade. O período de análise compreendido entre Setembro de 2014 e Março de 2015 corresponde a 151 dias (variações diárias) e 29 semanas (variações semanais). A componente *Size* indica o tamanho dos *itemsets* que ocorrem para cada entidade, em cada momento do tempo. Note-se que, para cada entidade, em cada momento do tempo ocorre apenas uma variação do preço de fecho e, portanto, o *Size* será sempre 1. Por fim, a componente *Item* identifica a variação que ocorre em determinada *sequenceID* e *eventID*. Esta componente assume a categorização explicada nas tabelas 3.3 e 3.4.

A tabela 3.5 apresenta a forma como são representados os dados, usando a discretização mais grossa, que serão, posteriormente, aplicados ao cSPADE. Verifica-se, por exemplo, que na empresa Adidas (*sequenceID* = 1), no momento 1, ocorre uma variação do preço de fecho entre -1% e 1% (categoria “E”), no momento 2, foi registada uma variação superior a 1% (categoria “A”) também na empresa Adidas.

sequenceID	eventID	Size	Item
1	1	1	E
1	2	1	A
1	3	1	E
(...)	(...)	(...)	(...)
2	1	1	E
2	2	1	A
(...)	(...)	(...)	(...)

Tabela 3.5: Representação dos dados

3.4 Experiências Realizadas

Após o pré-processamento dos dados, o próximo passo consiste na extração dos padrões sequenciais a partir dos dados financeiros em análise.

As experiências realizadas nesta secção podem ser divididas em três grupos. O primeiro grupo consiste na extração dos padrões sequenciais a partir do conjunto de dados. No segundo grupo será realizada uma análise de similaridade aos movimentos dos

preços das entidades. Por fim, no terceiro grupo será realizada, de forma individual, a extração de padrões sequenciais em cada empresa, bem como a respetiva análise de similaridade.

Em seguida será feita uma breve introdução às diferentes experiências que serão realizadas em todos os grupos.

3.4.1 Motivação

O primeiro grupo de experiências consiste na aplicação do algoritmo cSPADE aos dados em estudo a fim de encontrar os padrões sequenciais que caracterizem os movimentos de preços de fecho das dezanove entidades, durante os sete meses, e possibilitem a obtenção de conhecimento útil para o investidor.

Neste grupo, será realizada a análise de granularidade temporal a fim de avaliar o impacto do uso de dados diários e semanais nos resultados obtidos. Por outro lado, em todas as experiências realizadas neste grupo serão definidos dois tipos de discretização das variações dos preços. Na discretização mais grossa (ver tabela 3.3) são estabelecidas três categorias de variação dos preços os quais indicam aumentos, diminuições ou estabilizações dos mesmos. Contudo, poderá ser interessante explorar os aumentos e diminuições (superiores a 2% e inferiores a -2%), nomeadamente se estes correm com alguma frequência. Para tal na discretização mais fina são adicionadas mais duas categorias de variação que estudam as variações entre 1% e 2% e entre -2% e -1% (ver tabela 3.4). Adicionalmente, importa perceber o impacto que o uso de dados diários, que apresentam mais momentos de variação dos preços em relação aos dados semanais, ou o uso de mais categorias de variação (discretização mais fina), possam ter no desempenho do algoritmo, nomeadamente no tempo de processamento e no número de sequências frequentes.

Tal como referido anteriormente, a escolha do algoritmo cSPADE foi motivada pelo facto do mesmo permitir a incorporação de restrições adicionais que permitem focar a análise na informação que possa produzir mais conhecimento para o utilizador. Contudo, importa perceber a relevância que tem o uso destas restrições e para isso, neste grupo, serão realizadas experiências apenas com o uso mínimo de restrições e mais à

frente, na secção 3.4.2.3, serão incorporadas outras restrições. Neste sentido, além da implementação de um suporte mínimo, que é necessário para a descoberta de sequências frequentes, apenas será usada uma restrição que está relacionada com o tamanho máximo das sequências encontradas. Este tamanho máximo das sequências será estabelecido em sete, ou seja, cada sequência frequente encontrada terá no máximo sete variações dos preços de fecho. A justificação para tal deve-se ao facto de que as sequências longas não produzirem conhecimento útil para o utilizador, como demonstrado em *Wu et al.* (2014).

De seguida será realizada a descoberta de sequências frequentes, com apenas duas restrições: o tamanho máximo da sequência e o suporte mínimo. Contudo, a definição de um suporte mínimo é um processo complexo visto que, por um lado, um suporte mínimo elevado poderá conduzir à perda de informação relevante e, por outro lado, a definição de um suporte mínimo baixo produz um grande número de sequências que poderão não incorporar conhecimento importante. Neste sentido, serão testados diferentes limites de suporte a fim de perceber qual o nível que melhores resultados apresenta.

A segunda experiência a realizar no primeiro grupo passa pela implementação de duas restrições adicionais: *maxwin*, que corresponde à diferença temporal máxima entre dois quaisquer itens numa sequência, e *maxgap*, que limita a ocorrência máxima entre dois itens consecutivos numa sequência. Pretende-se responder a questões como: será que a definição de restrições adicionais permite a obtenção de melhores resultados? Que tipo de restrições poderão ser implementadas que melhor expliquem o comportamento dos dados em estudo?

A terceira e última experiência, neste primeiro grupo, consiste em, a partir dos resultados obtidos anteriormente, nomeadamente das definições que melhores resultados produzem, descobrir as sequências máximas. Vários autores, como Agrawal et al. (1995), defendem que a extração de padrões frequentes não deve ocorrer sobre todos os padrões frequentes, mas sim apenas sobre os padrões frequentes máximos.

No segundo grupo (secção 3.4.3) será realizada a análise de similaridade aos movimentos dos preços das diversas entidades incluídas neste estudo. A análise de similaridade consiste em verificar que entidades ou grupos de entidades apresentam um comportamento mais semelhante/dissemelhante durante o período em análise. Este tipo de análise pode fornecer conhecimento útil para o investidor. Note-se que, segundo a

teoria económica, a estratégia a adotar, por parte dos investidores, para reduzir o risco de uma carteira de investimentos, passa por apostar em ativos com comportamentos diferentes, ou seja, com movimentos de preços opostos. Assim, o objetivo desta análise prende-se com a possibilidade de descobrir semelhanças/dissemelhanças no comportamento destas entidades que possibilitem de obtenção de informação característica sobre os instrumentos financeiros em causa.

Note-se que, nos grupos anteriores, serão realizadas diversas experiências tendo como referência o conjunto das entidades em análise. Porém, no terceiro grupo (secção 3.4.4) será realizada, de forma individual, a extração de padrões sequenciais em cada uma das entidades em análise. O objetivo consiste na tentativa de fornecer algum conhecimento para o utilizador sobre cada entidade em estudo através dos padrões sequências descobertos e da análise de similaridade realizada na secção anterior.

3.4.2 Extração de Padrões Sequenciais

De seguida serão estudados os padrões sequenciais encontrados a partir do conjunto de dados. Através de várias experiências, pretende-se analisar o impacto da granularidade temporal dos dados e da granularidade de discretização das variações dos preços (secções 3.4.2.1 e 3.4.2.2), o efeito da aplicação de restrições na aprendizagem de sequências (secção 3.4.2.3) e, por fim, a extração de sequência frequentes máximas (secção 3.4.2.4).

3.4.2.1 Extração de sequências frequentes com dados diários

Nesta seção será realizada a extração de sequências frequentes com dados diários, usando apenas duas restrições: o suporte e o tamanho máximo das sequências. O objetivo consiste em avaliar o impacto da granularidade de discretização dos dados no processo de extração de sequências frequentes.

De referir que, com dados diários, em ambos os tipos de discretização, o período de análise compreendido entre Setembro de 2014 a Março de 2015 corresponde a 151 dias de variações relativas dos preços de fecho. Neste sentido, e tendo em conta que são analisadas 19 empresas/entidades, a base de dados com dados diários apresenta um total de 2.869 transações. Note-se que cada transação indica uma variação do preço de fecho numa empresa, num determinado momento do tempo (neste caso, em cada dia).

No que diz respeito à definição do suporte mínimo, serão testados diversos limites mínimos de suporte, nomeadamente: 40%, 50%, 60%, 70%, 80% e 90%. Com os diferentes níveis de suporte serão analisados os tempos de execução do algoritmo bem como o número de sequências encontradas.

A figura 3.1 apresenta o tempo de processamento do algoritmo e o número de sequências encontradas quando são estabelecidos diferentes limites de suporte. É possível verificar que à medida que o suporte mínimo aumenta o tempo de processamento do algoritmo diminui. Esta situação decorre do facto de o aumento do suporte mínimo exigido para que uma sequência seja considerada frequente conduzir a que muitas sequências não sejam frequentes, diminuindo o espaço de procura do algoritmo e, assim, o tempo de processamento para encontrar as sequências frequentes diminui.

Relativamente ao número de sequências frequentes encontradas, importa esclarecer que ao estabelecer o tamanho máximo estão a ser excluídas um grande número de sequências. Contudo, a implementação desta restrição é um procedimento importante de forma a limitar a descoberta de sequências longas que não produzem conhecimento útil no âmbito desta análise. Do ponto de vista do investidor, poderá não ser útil a extração de sequências demasiado longas pelo facto de não produzirem informação relevante e de, eventualmente, dificultarem a tomada de decisão.

Neste sentido, verifica-se que a alteração do suporte mínimo não modifica o número de sequências encontradas. Tal deve-se ao facto de as sequências frequentes encontradas com tamanho máximo de 7 itens apresentarem um suporte elevado sendo, assim, extraídas em qualquer um dos limites que são apresentados. Assim, o número de sequências encontradas com dados diários, na discretização mais grossa, e em qualquer um dos suportes mínimos estabelecidos é de 3.279.

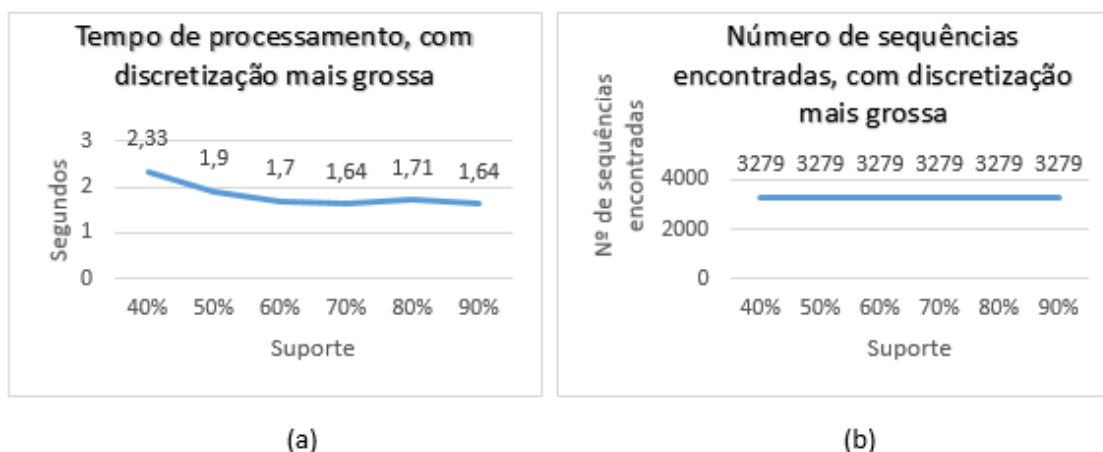


Figura 3.1: Tempo de processamento (a) e número de sequências encontradas (b), com dados diários, na discretização mais grossa, com diferentes limites de suporte

Com a discretização mais fina nas variações verifica-se que o tempo de execução aumenta consideravelmente (figura 3.2). Existem, contudo, dois aspetos que não se alteram quando comparado com a discretização mais grossa. O primeiro diz respeito ao facto do tempo de execução diminuir à medida que o limite mínimo de suporte aumenta, o segundo está relacionado com o número de sequências encontradas. Neste último caso, a limitação do tamanho das sequências encontradas a 7, revela que uma grande quantidade dessas sequências apresenta um suporte elevado. Na configuração dos diferentes tipos de suporte, o algoritmo encontra 97.655 sequências que apresentam um suporte mínimo de 70%, porém destas cerca de 108 não atingem o suporte mínimo de 80% e cerca de 2.092 não apresentam um suporte de 90%.

Importa, contudo, referir que o número de sequências encontradas com discretização mais fina, mesmo com um suporte elevado (90%), é intratável. Neste sentido, a implementação de restrições adicionais que, por um lado limitem o espaço de procura do algoritmo, por outro lado, foquem a análise na questão em estudo, é um procedimento de extrema importância na extração de padrões sequenciais. Este procedimento será tratado no ponto 3.4.2.3.

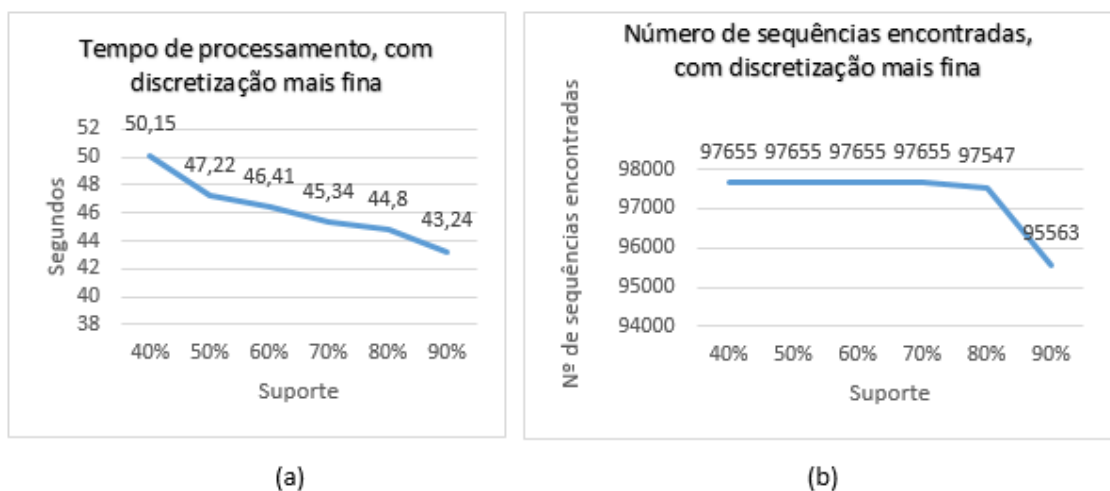


Figura 3.2: Tempo de processamento (a) e número de sequências encontradas (b), com dados diários, na discretização mais fina, com diferentes limites de suporte

3.4.2.2 Extração de sequências frequentes com dados semanais

As experiências realizadas nesta seção são semelhantes às do ponto anterior, com exceção de que agora são usados os dados semanais. Visto isto, o período compreendido entre Setembro de 2014 e Março de 2015 compreende um total de 29 semanas de variações, sendo que para as 19 entidades foram realizadas 551 transações. Note-se que, à semelhança do efetuado com os dados diários, será mantida a restrição do comprimento máximo das sequências em 7, assim como, serão testados diversos limites de suporte (40%, 50%, 60%, 70%, 80% e 90%).

O tempo de processamento do algoritmo, com dados semanais na discretização mais grossa, varia entre 1,68 segundo (suporte 40%) e 0,94 segundos (suporte 90%) (figura 3.3). Ao ser comparado com os tempos de processamento com dados diários e na discretização mais grossa, verifica-se que com dados semanais, o algoritmo é mais eficiente. Tal pode ser justificado com o facto de a base de dados com variações semanais (551 transações) ser mais reduzida do que com variações diárias (2.869 transações). Por outro lado, verifica-se que a alteração do suporte conduz, também, à alteração do número

de sequências descobertas. Com um suporte de 40%, o algoritmo encontra cerca de 3.272 sequências, um número muito semelhante ao verificado com dados diários (3.279), contudo, ao contrário dos dados diários, estas sequências não apresentam um suporte tão elevado, visto que à medida que o mesmo vai aumentando o número de sequências frequentes vai diminuindo.

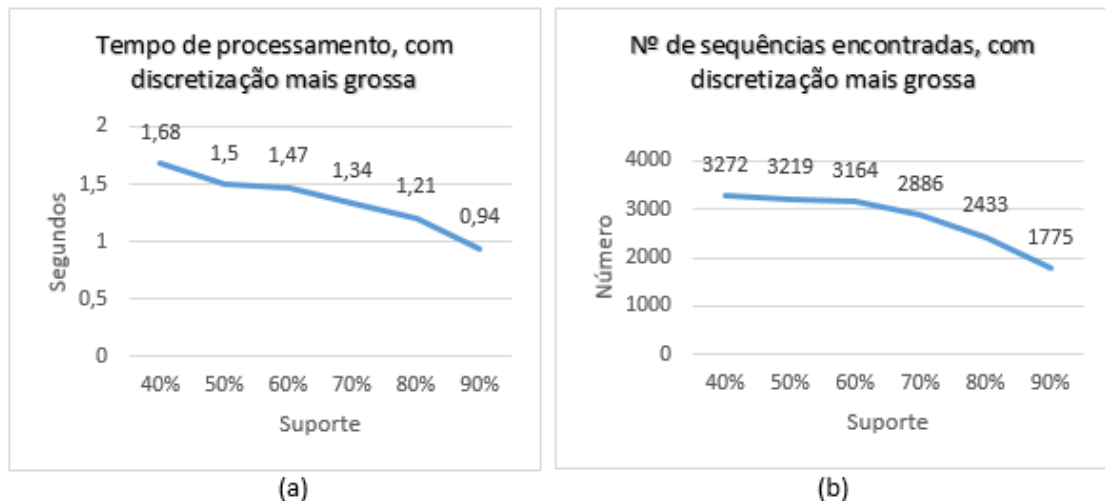


Figura 3.3: Tempo de processamento (a) e número de sequências encontradas (b), com dados semanais, na discretização mais grossa, com diferentes limites de suporte

Com a discretização mais fina, à semelhança do verificado com dados diários, tanto o tempo de processamento como o número de sequências encontradas aumenta em relação à discretização mais grossa. Note-se, contudo, que a partir do suporte mínimo de 80%, o algoritmo apresenta um melhor desempenho na discretização mais fina, com um tempo de execução mais baixo e um menor número de sequências encontradas.

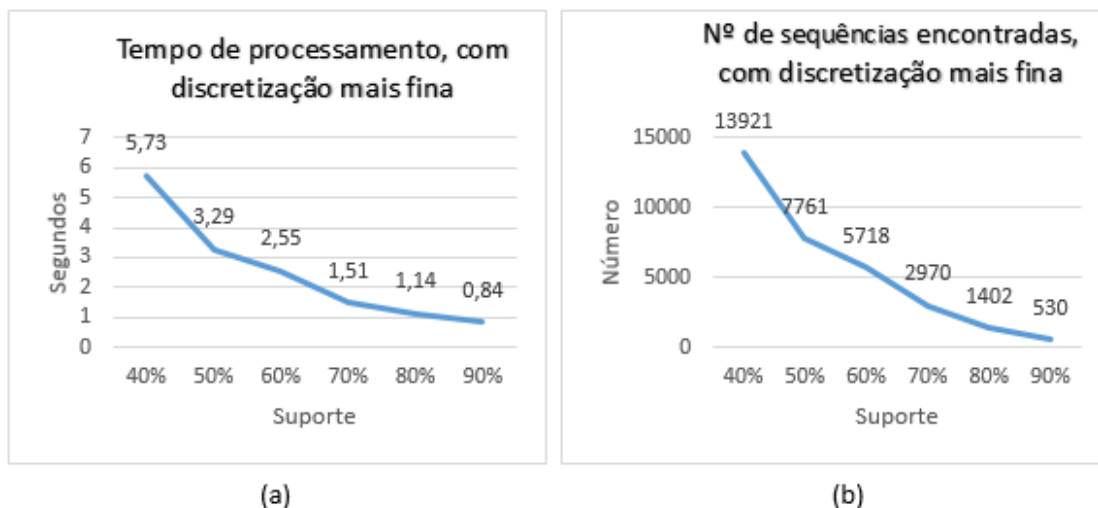


Figura 3.4: Tempo de processamento (a) e número de sequências encontradas (b), com dados semanais, na discretização mais fina, com diferentes limites de suporte

A implementação de dois tipos de discretização das variações permite compreender de que forma uma discretização mais fina pode alterar os resultados obtidos. Verifica-se que, tanto com dados diários como com dados semanais, o tempo de processamento e o número de sequências encontradas é maior quando os dados apresentam uma discretização mais fina, à exceção do ocorrido quando o suporte mínimo é igual ou superior a 80%, no caso dos dados semanais.

Adicionalmente, o acréscimo de duas novas categorias (discretização mais fina) pode produzir outro tipo de informação, nomeadamente a relativa à estabilidade/instabilidade dos movimentos dos preços das ações em estudo. A figura 3.5 apresenta a frequência da ocorrência dos itens nas variações das sequências encontradas, para os dois tipos de discretização, nos dois tipos de dados, usando um suporte mínimo de 80%.

Com dados diários, na discretização mais grossa, verificou-se anteriormente, que o algoritmo encontra o mesmo número de sequências em todos os limiares de suporte utilizados e, portanto, as sequências frequentes encontradas com um suporte de 40% apresentam uma frequência elevada (igual ou superior a 90%). Neste sentido, pela figura 3.5 (a), verifica-se que, para este caso, a distribuição da frequência dos itens pelas sequências é equitativa e, assim, as categorias de variação ocorrem, sensivelmente, o mesmo número de vezes. Ao transitar para uma discretização mais fina, os resultados não apresentam grandes alterações (ver figura 3.5 (b)). Apesar de serem diferenças muito reduzidas, a categoria “E”

(variações do preço entre -1% e 1%) é a que mais ocorre no total de variações (20,01%), seguindo-se a categoria “B” e “C”, com ocorrência de aproximadamente 20% em cada uma. Face às diferenças muito reduzidas na ocorrência dos itens, não é possível qualquer conclusão sobre o comportamento estável/instável dos preços.

Com dados semanais as conclusões alteram-se ligeiramente. Na discretização mais grossa, o item “A”, relativo a variações positivas do preço de fecho, é o que mais ocorre (34,4% no total de variações). Esta situação reflete um comportamento dos preços positivo durante o período em análise. Adicionalmente, como os dados sugerem um comportamento positivo dos preços, poderá ser interessante analisar a extensão destas variações positivas nos dados, nomeadamente através da implementação de duas categorias adicionais que irão explorar as variações entre -2% e -1% (categoria “C”) e entre 1% e 2% (categoria “B”). Os resultados obtidos na discretização mais fina, com dados semanais, permitem a verificação de uma maior frequência das variações mais elevadas, com os itens “A” (crescimento do preço superior a 2%) e “D” (decréscimo do preço superior a 2%) a ocorrerem em 33,4% e 29,9% das variações, respetivamente. Estas categorias indicam variações do preço de fecho superiores a 2% (“A”) e inferiores a -2% (“D”), o que leva a concluir que, com dados semanais, os movimentos dos preços deste grupo de ações apresentam alguma volatilidade durante o período em análise. Seria, contudo, de esperar maiores alterações nos dados semanais, visto que uma variação semanal pode representar a agregação das variações diárias.

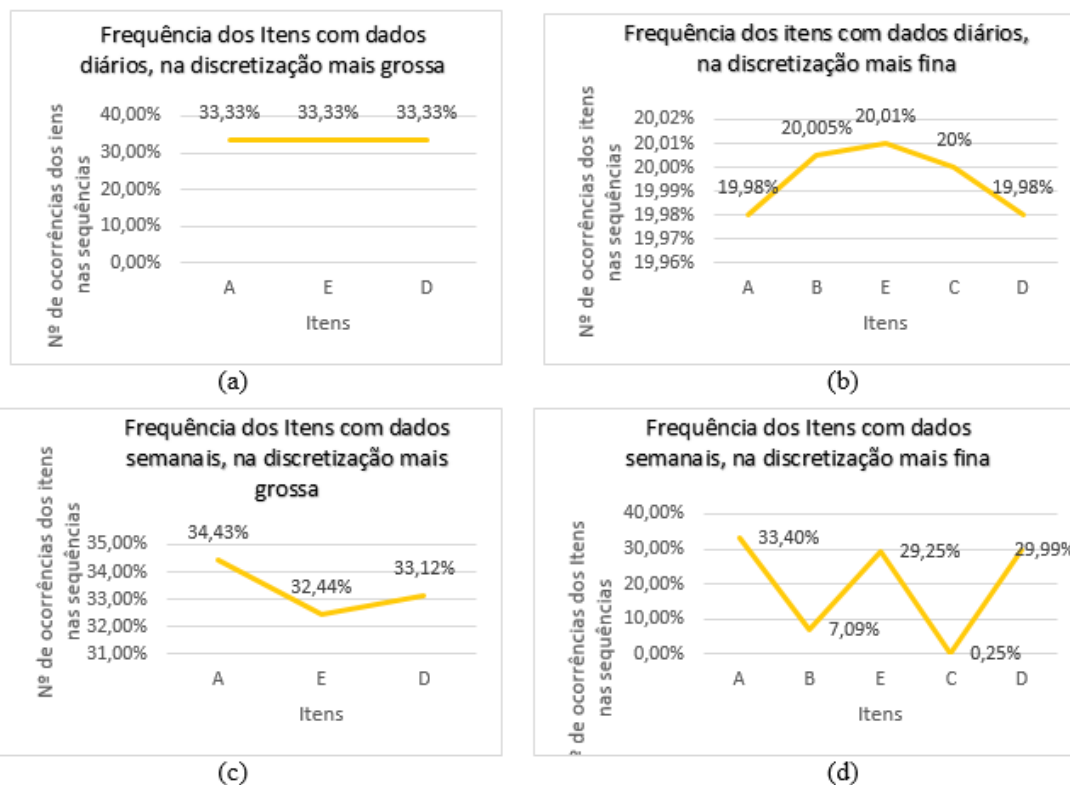


Figura 3.5: Frequência dos itens com dados diários na discretização mais grossa (a) e mais fina (b), e com dados semanais na discretização mais grossa (c) e mais fina (d)

Resumo das experiências realizadas nas secções 3.4.2.1 e 3.4.2.2

Nesta primeira parte de experiências realizadas no grupo 1, foi proposto o estudo da granularidade temporal dos dados por forma a analisar o impacto, nos resultados obtidos, da alteração de dados com variações diárias para dados com variações semanais. Verificou-se que o algoritmo apresenta um melhor tempo de execução com dados semanais, o qual pode ser motivado pelo facto de a base de dados ter menos transações do que a com variações diárias, e que determina um menor espaço de procura para o algoritmo (diminuição do tempo de execução). Por outro lado foi proposta, para esta primeira parte, a análise do impacto da granularidade de discretização dos intervalos de variação nos resultados obtidos. Verificou-se que a adição de duas novas categorias aumentava o número de sequências frequentes encontradas e, conseqüentemente, tornava o algoritmo ligeiramente mais lento, à exceção de, com dados semanais, quando o suporte mínimo era de 80%, em que, nesse caso, o algoritmo apresenta um tempo de execução

menor na discretização mais fina do que na discretização mais grossa. Por último, a adição de duas novas categorias permite um estudo mais detalhado acerca da estabilidade/instabilidade dos preços no período em análise, o que pode ser uma importante ferramenta de decisão para os investidores. Verificou-se que os dados diários apresentam algumas limitações pelo facto de incluírem um grande número de sequências frequentes com um suporte elevado e, portanto, uma distribuição da ocorrência dos itens muito equitativa.

Nas variações semanais, na discretização mais grossa (“A”, “E” e “D”) os resultados indicam um comportamento positivo nos preços (o item “A” ocorre com muita frequência). Para uma melhor compreensão deste comportamento positivo dos preços optou-se por explorar as variações superiores a 1% e inferiores a -1% através da adição de duas novas categorias (categoria “B” para variações entre 1% e 2% e categoria “C” para variações entre -2% e -1%). Assim, verificou-se que com dados semanais, na discretização mais fina, os preços de fecho do grupo de ações, durante o período em análise, apresentam alguma variabilidade, com os itens “A” (variações superiores a 2%) e “D” (variações inferiores a -2%) a ocorrerem em 33,4% e 29,9% das variações, respetivamente.

Tendo em conta que, não é conveniente usar um suporte demasiado baixo, devido ao grande número de sequências que são descobertas, e um suporte muito elevado, que possa conduzir à perda de informação útil, optou-se por estabelecer um suporte mínimo de 80% nas próximas experiências a realizar.

3.4.2.3 Extração de sequências frequentes com restrições

Nas experiências anteriores verificou-se que a implementação de apenas duas restrições (suporte e tamanho máximo das sequências) na procura por padrões sequenciais conduz a um elevado número de sequências frequentes descobertas. Assim sendo, e tendo em conta que é um procedimento defendido por diversos investigadores, o próximo passo consiste na implementação de um conjunto de restrições adicionais que permitam focar a análise no objeto em estudo e obter conhecimento útil.

O algoritmo cSPADE permite a definição de um conjunto vasto de restrições, nomeadamente: a limitação do comprimento das sequências; a implementação de restrições de gap máximo ou mínimo que irão limitar o tempo máximo/mínimo de ocorrência dos elementos consecutivos numa sequência; e ainda, a definição de janelas temporais na descoberta de sequências, o que determina o intervalo máximo de ocorrência de uma sequência frequente. Visto isto, além do uso de um suporte mínimo de 80% e do estabelecimento de um comprimento máximo das sequências em 7 itens, serão usadas as restrições de *maxwin* e *maxgap*. A restrição *maxwin* permite estabelecer a diferença temporal máxima entre quaisquer dois elementos de uma sequência e será definida em 4, ou seja, a diferença temporal máxima entre dois itens de uma mesma sequência é de 4 dias/semanas. Por exemplo, na sequência $\langle A \rightarrow E \rightarrow A \rangle$, para dados diários, com discretização mais grossa, com determinada probabilidade, ocorre uma variação do preço de fecho superior a 1%, seguindo-se uma variação entre -1% e 1% e, posteriormente, uma variação do preço superior a 1%, no período máximo de 4 dias.

A restrição *maxgap* limita a diferença temporal máxima entre dois itens consecutivos numa sequência e será definida em 1, isto é, a ocorrência máxima de dois itens consecutivos numa sequência é 1 dia/semana. Por exemplo, na sequência $\langle B \rightarrow A \rightarrow E \rightarrow C \rangle$, com dados semanais e na discretização mais fina, a variação do preço de fecho superior a 2% (categoria “A”) ocorre no máximo uma semana depois da variação do preço entre 1% e 2% (categoria “B”). Estas restrições serão, assim, adicionadas ao algoritmo na tentativa de melhorar os resultados anteriormente obtidos.

Pela figura 3.6, verifica-se que a adição das restrições *maxwin* e *maxgap* tem impactos diferentes consoante o tipo de dados usados. Com dados diários, na discretização mais grossa, o algoritmo apresenta um tempo de execução de 1,84 segundos no qual encontra cerca de 3.382 sequências (1,71 segundos e 3.279 sequências, sem as restrições adicionais). Na discretização mais fina, o tempo de execução aumenta consideravelmente, bem como o número de sequências encontradas (2,75 segundos e 6.022 sequências) comparativamente com a discretização mais grossa, contudo, apresenta um bom desempenho em relação às experiências sem restrições (44,8 segundos e 97.547 sequências). O desempenho menos favorável do algoritmo, com a discretização mais fina em relação à discretização mais grossa, pode estar relacionada com o facto de os dados

apresentarem mais categorias e aumentar, assim, o espaço de procura. Contudo, verifica-se uma melhoria do desempenho (tempo de execução de número de sequências encontradas) com a adição das restrições.

Com dados semanais são obtidos resultados diferentes. A adição das restrições conduz, na discretização mais grossa, a uma redução significativa do tempo de processamento (0,69 segundos) e do número de sequências descobertas (1.104) em relação ao verificado nas experiências sem as restrições (1,21 segundos e 2.433 sequências encontradas). Na discretização mais fina o tempo de execução e o número de sequências são ainda menores.

Assim, concluímos que a adição de duas novas restrições permite, tanto com dados diários como com dados semanais, diminuir o espaço de procura do algoritmo o que torna o processo de descoberta de sequências mais eficiente.

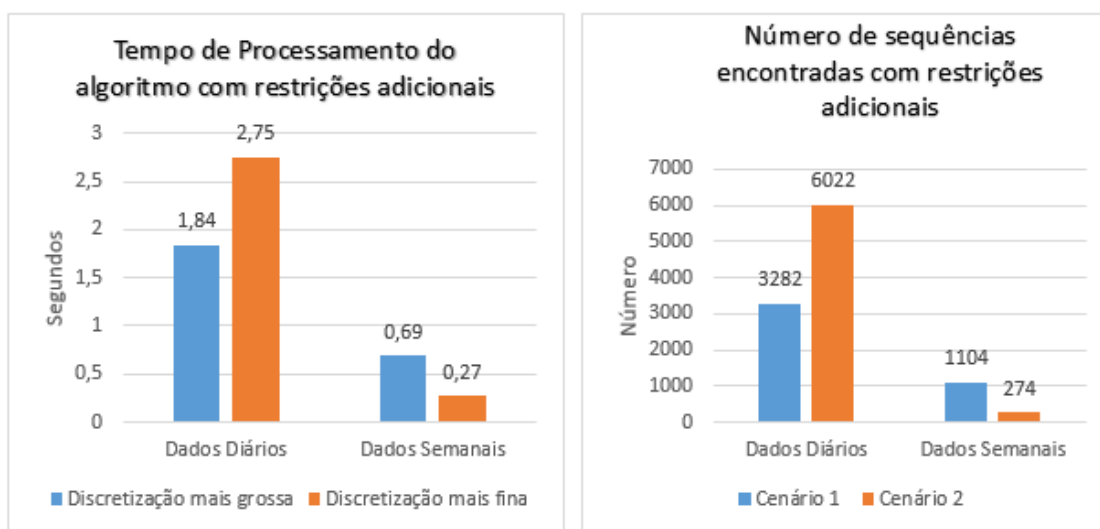


Figura 3.6: Tempo de processamento (a) e número de sequências encontradas (b) com restrições adicionais

Nas experiências anteriores (sem as restrições adicionais) verificou-se que a distribuição da frequência dos itens nos dados diários era muito equitativa, enquanto com dados semanais era possível observar alguma variabilidade nos movimentos das ações, com os itens “A” (variações nos preços superiores a 2%) e “D” (variações nos preços inferiores a -2%) a ocorrerem com mais frequência que os restantes.

Com a adição de restrições, na discretização mais grossa, verifica-se que a distribuição da ocorrência dos itens é muito semelhante nos dados diários, contudo, no caso dos dados semanais, os itens “A” e “D” ocorrem com mais frequência que o item “E”, tal como nas experiências sem as restrições adicionais (figura 3.7). No caso dos dados diários, na discretização mais grossa verifica-se uma maior ocorrência do item “E” em relação aos restantes, contudo, as diferenças são muito reduzidas (figura 3.7 (a)). Na discretização mais fina, as variações dos preços entre -1% e 1% são as que mais ocorrem (29,9%), tal como acontecia na experiência sem restrições adicionais, contudo, neste caso, as diferenças são maiores em relação à frequência da ocorrência dos outros itens.

No caso dos dados semanais, à semelhança do verificado nas experiências anteriores, os movimentos dos preços indicam um comportamento positivo, ou seja, uma maior ocorrência de aumentos superiores a 2% (item “A”) nos itens das sequências frequentes. É, ainda, interessante verificar que os preços evidenciam uma maior tendência para variações semanais superiores a 2% (item “A”), inferiores a -2% (item “D”) e estabilizações dos preços entre -1% e 1% (item “E”), do que para variações entre -2% e -1% (item “C”) e entre 1% e 2% (item “B”).

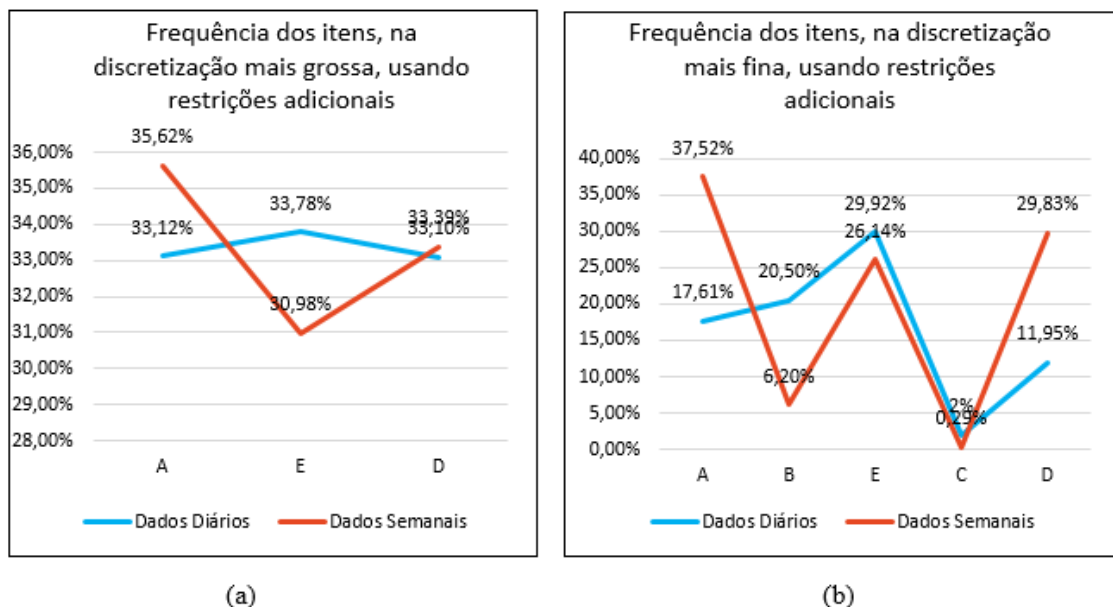


Figura 3.7: Frequência dos itens com o uso de restrições adicionais, na discretização mais grossa (a) e frequência dos itens com o uso de restrições adicionais, na discretização mais fina (b)

Na tabela 3.6 são apresentadas algumas das sequências frequentes descobertas, na discretização mais fina, com dados diários e com o uso das restrições adicionais. Por exemplo, a sequência $\langle A \rightarrow E \rightarrow A \rightarrow A \rangle$ indica que, com 84% de frequência, num intervalo temporal máximo de 4 dias, ocorre uma variação no preço de fecho superior a 2%, seguindo-se uma variação entre -1% e 1%, e em seguida ocorrem duas variações superiores a 2%. A sequência $\langle B \rightarrow E \rightarrow A \rightarrow A \rangle$ ocorre com 95% de frequência, onde surge uma variação do preço de fecho entre 1% e 2%, seguindo-se uma variação entre -1% e 1%, uma variação superior a 2% e uma outra variação superior a 2%, no prazo máximo de 4 dias.

Na tabela 3.7 são apresentadas algumas das sequências frequentes com dados semanais. A sequência $\langle D \rightarrow E \rightarrow A \rightarrow E \rangle$ ocorre com 89% de frequência e descreve uma variação do preço de fecho inferior a -2%, seguindo-se uma variação entre -1% e 1%, uma variação superior a 2% e, posteriormente, o preço de fecho apresenta uma variação entre -1% e 1%, durante um intervalo de 4 semanas.

Sequência	Suporte
$\langle B \rightarrow E \rightarrow A \rightarrow A \rangle$	0.9473684
$\langle E \rightarrow E \rightarrow C \rightarrow E \rightarrow A \rightarrow A \rangle$	0.9473684
$\langle A \rightarrow E \rightarrow A \rightarrow A \rangle$	0.8421053
$\langle D \rightarrow B \rightarrow A \rightarrow A \rangle$	0.7894737
$\langle E \rightarrow A \rightarrow A \rightarrow A \rangle$	0.7894737

Tabela 3.6: Sequências frequentes, na discretização mais fina, com dados diários

Sequência	Suporte
$\langle D \rightarrow E \rightarrow A \rightarrow E \rangle$	0.8947368
$\langle D \rightarrow A \rightarrow E \rightarrow E \rangle$	0.7894737
$\langle A \rightarrow E \rightarrow D \rightarrow E \rangle$	0.7894737
$\langle D \rightarrow D \rightarrow A \rightarrow E \rangle$	0.7894737
$\langle D \rightarrow E \rightarrow A \rightarrow A \rightarrow B \rangle$	0.7894737
$\langle A \rightarrow A \rightarrow A \rightarrow A \rightarrow A \rightarrow E \rightarrow A \rangle$	0.7894737

Tabela 3.7: Sequências frequentes, na discretização mais fina, com dados semanais

De acordo com os resultados obtidos, conclui-se que o uso de restrições adicionais no processo de descoberta de sequências reduz significativamente o espaço de procura e, consequentemente, o número de sequências encontradas, o que torna o algoritmo mais eficiente. Por outro lado, o uso destas restrições permite focar a análise no objetivo da

dissertação, a procura por conhecimento útil que facilite o processo de tomada de decisão dos investidores. Tendo em conta que as restrições implementadas dizem respeito a limitações máximas temporais, os resultados obtidos fornecem informação útil sobre a ocorrência máxima dos itens de uma sequência, a qual pode ser usada estrategicamente nas decisões de investimento.

Neste sentido, as próximas experiências a realizar irão incluir estas restrições. Adicionalmente, o uso da discretização com cinco categorias de variação permite uma melhor exploração das variações de preços que não é possível com apenas três categorias e, assim, além do uso das restrições adicionais, as próximas experiências serão realizadas tendo em conta a discretização mais fina.

3.4.2.4 Extração de sequências frequentes máximas

Vários autores defendem que a extração de padrões frequentes não deve ocorrer sobre todos os padrões frequentes, mas sim sobre os máximos. Uma sequência é máxima se é frequente e nenhuma das suas supersequências é frequente, ou seja, uma sequência é máxima se não está contida em mais nenhuma sequência frequente (ver secção 2.3).

Existem algoritmos que, aquando da contagem das sequências frequentes, contam sequências não máximas, mas que, posteriormente, na fase de poda, irão eliminá-las. Um exemplo destes algoritmos, que são designados de algoritmos *count-all*, é o *AprioriAll*.

O algoritmo *cSPADE* contabiliza também as sequências não máximas e, assim, é importante incorporar uma fase na qual, após a descoberta das sequências frequentes estas sejam eliminadas e, posteriormente, sejam guardadas apenas as sequências frequentes máximas. Neste sentido, iremos proceder à extração das sequências frequentes máximas, nos dados diários e nos dados semanais, com a discretização mais fina, usando um suporte mínimo de 80% e as restrições anteriormente indicadas.

O número de sequências frequentes encontradas anteriormente com dados diários, na discretização mais fina, foi de 6.022, contudo, destas, apenas 3.137 são máximas. Com dados semanais, obtemos um total de 274 sequências frequentes nas experiências

anteriores, sendo que apenas 68 são máximas. As sequências frequentes máximas encontradas com dados diários e com dados semanais podem ser verificadas nas tabelas 6.1 e 6.2 dos anexos.

3.4.3 Análise de Similaridade

As séries temporais consistem em grandes quantidades de dados ordenados temporalmente. A capacidade de encontrar séries temporais que são “similares” a outras ou a descoberta de grupos de séries temporais similares entre si pode produzir informação interessante, nomeadamente quando as séries temporais reproduzem os movimentos de preços de um conjunto de entidades.

Ao estudar séries temporais que traduzem as variações do preço de fecho de determinadas ações num determinado espaço de tempo, é possível, através de medidas de similaridade, obter relações interessantes entre essas variações. Esta análise produz informação útil para o investidor na medida em que torna possível a identificação das ações/grupos de ações que apresentam maior/menor similaridade, o que permite a diversificação do risco da carteira através do investimento em ações dissemelhantes. Esta hipótese designa a Teoria da Eficiência de Markowitz (*Markowitz, H., 1952*).

Neste sentido, será realizada a análise de similaridade das variações do preço de fecho das ações, decorridas durante o período de Setembro de 2014 a Março de 2015, relativas às 19 empresas/entidades em estudo. O objetivo passa por descobrir relações no comportamento destas ações com vista à obtenção de conhecimento útil para o investidor. O método usado para encontrar a similaridade entre as sequências é o método *Jaccard*. A medida de similaridade de *Jaccard* entre dois conjuntos de dados/sequências consiste no rácio entre o número de elementos em comum e o número de elementos que pertence pelo menos a uma das sequências tal que:

$$S(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

Por exemplo, considerando as duas sequências $X = \langle A \rightarrow B \rightarrow B \rightarrow C \rangle$ e $Y = \langle A \rightarrow B \rightarrow D \rangle$. $X \cap Y$ é dado por $\langle A \rightarrow B \rangle$ e $X \cup Y$ é $\langle A \rightarrow B \rightarrow C \rightarrow D \rangle$. Assim, a similaridade entre X e Y é 0,5. Esta medida é uma comparação cronológica dos movimentos dos preços das entidades, ou seja, por exemplo, no caso dos dados diários, é comparada a variação do preço de uma ação no dia 1 com a variação do preço de outra ação também no dia 1, e assim sucessivamente.

Em seguida será realizada uma análise mais detalhada à similaridade em cada tipo de dados, usando a discretização mais fina.

3.4.3.1 Análise de similaridade com dados diários

Na tabela 3.8 é apresentada a similaridade entre as empresas/entidades em estudo, com variações diárias do preço de fecho, na discretização mais fina. As entidades que apresentam uma maior similaridade são a Basf e Siemens, Allianz e Linde, Allianz e Siemens, BMW e Daimler, Fresenius e Siemens, entre outras. O facto de estas empresas apresentarem uma similaridade elevada, significa que durante o período em análise, as mesmas apresentaram comportamentos, ao nível da variação diária do preço de fecho, muito semelhantes.

Por outro lado, ao analisar a dissemelhança ou menor similaridade verifica-se que entre as 19 entidades em análise, a que menos se assemelha ao comportamento das restantes entidades é o fundo de investimentos USA Oil Fund.

Note-se que a inclusão dos preços do fundo de investimento USA Oil Fund tinha como objetivo analisar possíveis relações no comportamento dos preços de instrumentos financeiros norte americanos e europeus. Nesta experiência verifica-se que os

movimentos dos preços do USA Oil Fund apresentam algumas dissemelhanças com as restantes entidades.

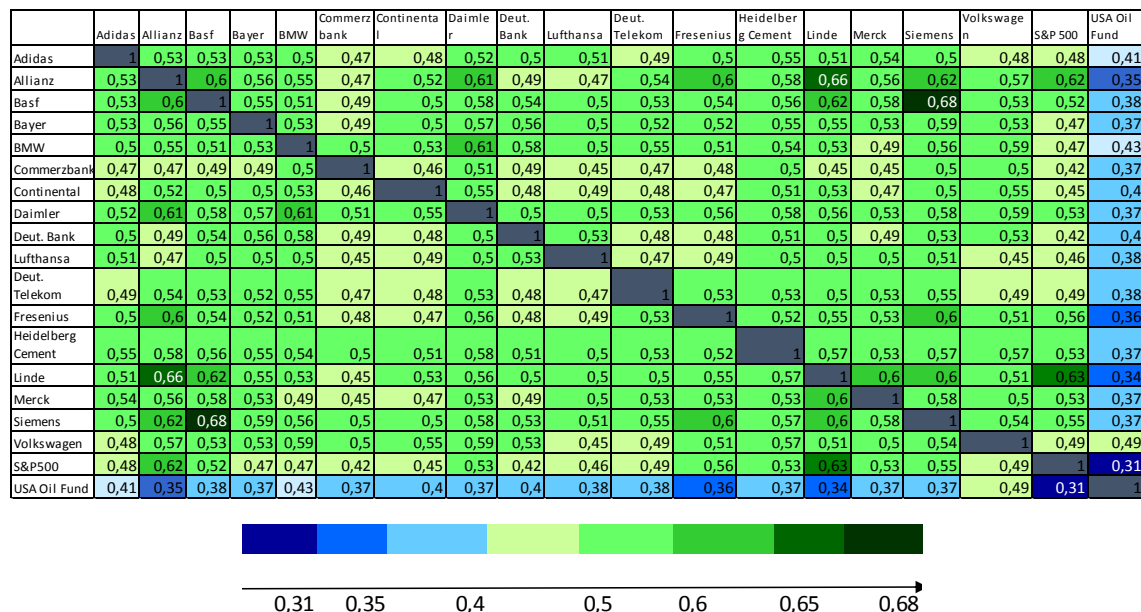


Tabela 3.8: Similaridade das entidades com dados diários, na discretização mais fina

3.4.3.2 Análise de similaridade com dados semanais

Com os dados semanais, alguns aspetos verificados anteriormente alteram-se. Neste caso, surgem como as empresas que mais similaridade apresentam a Daimler e a Volkswagen, o que com os dados diários não se verifica, assim como outras relações entre outras entidades (tabela 3.9). Por outro lado, a análise de similaridade permite verificar comportamentos, na variação do preço de fecho, semelhantes entre empresas do mesmo setor. Neste caso, é possível verificar uma maior similaridade entre empresas do setor automóvel, como a BMW, Volkswagen, Continental e a Daimler.

As relações que mais dissemelhanças apresentam, quando usados os dados semanais, ocorrem no fundo de investimentos USA Oil Fund, tal como com os dados diários, mas também com o índice S&P500 e entre outras empresas como Basf e

Lufthansa, Daimler e Linde, Volkswagen e Linde. Ao contrário do verificado nos dados diários, em que a maior dissimilaridade ocorre apenas com o fundo de investimentos USA Oil Fund e as outras empresas, quando são usados os dados semanais verifica-se maiores dissimilaridades entre as entidades.

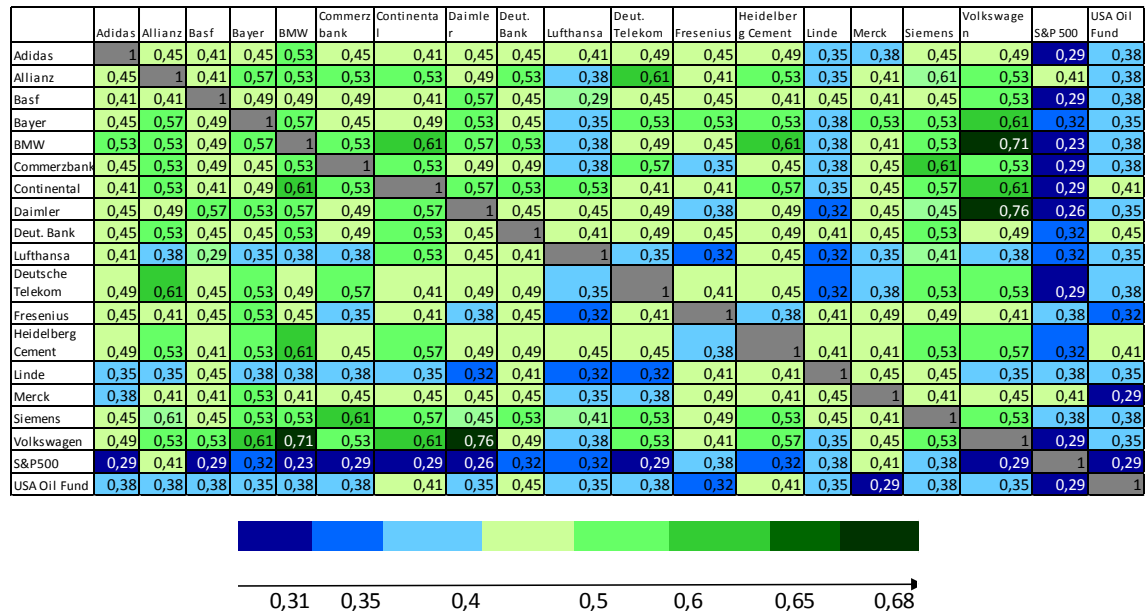


Tabela 3.9: Similaridade das entidades com dados semanais, na discretização mais fina

Em suma, esta análise produz informações de grande utilidade, nomeadamente no processo de tomada de decisão em investimentos. Segundo a teoria económica, um investidor, a fim de reduzir o risco, deve investir na diversificação da carteira (ações que apresentam comportamentos diferentes), neste sentido a análise à dissimilaridade das empresas pode ajudar na escolha de em quais investir.

Face ao exposto, conclui-se que a granularidade temporal dos dados pode ter um impacto relevante nos resultados obtidos. Em primeiro lugar, verificam-se maiores dissimilaridades com dados semanais (muitas similaridades abaixo dos 0,29) do que com dados diários (a maior dissimilaridade é de 0,31). Uma outra evidência deste estudo é que o uso de dados semanais permitiu a verificação de similaridades elevadas em empresas do mesmo setor de atividade, nomeadamente do setor automóvel, o que não foi possível com dados diários.

Por último, lembrando que a escolha da inclusão do índice S&P500 e do fundo de investimentos USA Oil Fund nesta análise tinha como objetivo a tentativa de descoberta de relações entre o comportamento de instrumentos financeiros norte-americanos e europeus, a análise de similaridades permitiu concluir que, durante o período em análise, o índice S&P500 e o fundo de investimento USA Oil Fund apresentam variações dos preços muito dissemelhantes em relação às restantes entidades em estudo. Importa, contudo, esclarecer que esta análise de similaridade é focada nas variações que ocorrem no mesmo período temporal, ou seja, é feita a comparação da variação dos preços de duas entidades ocorridos no mesmo dia ou semana. Neste sentido, o facto de existirem dissemelhanças nos preços não invalida que possam haver empresas que influenciam o comportamento dos preços de outras, com as variações a ocorrerem *à posteriori*. Por exemplo, o comportamento dos preços do fundo de investimentos USA Oil Fund pode, eventualmente, influenciar os movimentos de preços futuros das outras ações, contudo, tendo em conta a presente análise, com a medida de similaridade usada, e que compara os movimentos de preços ocorridos no mesmo instante temporal, verifica-se que o USA Oil Fund apresenta dissemelhanças com as restantes ações.

3.4.3 Estudo individualizado dos padrões sequenciais

As experiências realizadas anteriormente com a extração de padrões sequenciais, as alterações dos parâmetros do algoritmo (restrições) e exploração da granularidade temporal dos dados, foram efetuadas no conjunto dos dados relativos a todas as entidades em análise.

Nesta secção pretende-se efetuar uma análise mais individualizada a cada entidade presente no estudo. Para tal, é realizada a extração de sequências frequentes nos movimentos de preços de cada entidade. À semelhança do efetuado anteriormente, será usado o algoritmo cSPADE para a descoberta das sequências frequentes nas variações semanais, com um suporte mínimo de 80%. As restrições implementadas consistem no tamanho máximo da sequência em quatro variações e no maxgap de uma semana, ou seja,

a diferença temporal máxima ocorrida entre dois itens consecutivos numa sequência é de uma semana. Adicionalmente, pretende-se complementar esta análise com os resultados obtidos no estudo da similaridade (secção 3.4.3.2).

Adidas

Com dados diários, os movimentos dos preços de fecho das ações da Adidas apresentam maiores similaridades com a Heidelberg Cement (0,55) e com Merck (0,54) e uma maior dissemelhança com o índice USA Oil Fund (0,41). Com dados semanais, verifica-se uma maior semelhança com BMW (0,53) e menos semelhança com o índice S&P500 (0,29), a empresa Linde (0,35), o USA Oil Fund (0,38) e a com empresa Merck (0,38).

Foram encontradas 26 sequências frequentes máximas nos movimentos de preços da Adidas, durante o período em análise (ver tabela 6.3 dos anexos). Com um suporte superior a 80%, duas das sequências descobertas são: $\langle C \rightarrow D \rightarrow B \rightarrow E \rangle$ e $\langle A \rightarrow B \rightarrow A \rightarrow A \rangle$. A primeira indica a ocorrência de uma variação do preço de fecho entre -2% e -1%, seguindo-se uma variação inferior a -2%, um aumento do preço entre 1% e 2% e uma estabilização entre -1% e 1%. Esta informação pode ser interessante do ponto de vista do investidor uma vez que a mesma pode ser utilizada no apoio à decisão. Na segunda sequência verifica-se que, com frequência superior a 80%, após a ocorrência uma variação dos preços superior a 2%, uma estabilização entre -1% e 1% e um aumento superior a 2%, segue-se novamente um aumento do preço superior a 2%.

Allianz

As variações dos preços de fecho da Allianz, tanto no caso dos dados diários como com dados semanais, apresenta uma similaridade reduzida com o fundo de investimentos USA Oil Fund. Com dados diários verificam-se similaridades relevantes com a Linde (0,66), a Siemens (0,62), a Daimler (0,61), a Basf (0,6) e com a Fresenius (0,6). E com dados semanais, a similaridade mais elevada ocorre com a Deutsche Telekom (0,61) e a Siemens (0,61).

Nas variações dos preços da Allianz foram encontradas cerca de 26 sequências máximas frequentes as quais indiciam um comportamento positivo dos preços. As

variações dos preços superiores a 2% e entre -1% e 1% são as que registam um maior número de ocorrências nas sequências (ver figura 6.2 dos anexos).

BASF

A análise de similaridade realizada na secção anterior indica que as empresas que apresentam mais semelhanças com a BASF, relativamente às variações diárias dos preços de fecho, são a Siemens (0,68) e a Linde (0,62), por outro lado, o USA Oil Fund (0,38) é a entidade mais dissemelhante da BASF. Com variações semanais, o USA Oil Fund (0,38) continua a apresentar grandes dissemelhanças, assim como o índice S&P500 (0,29) e a Lufthansa (0,29).

Durante o período de Setembro de 2014 a Março de 2015, o algoritmo encontra cerca de 24 sequências máximas frequentes. Verifica-se que neste período, os preços de fecho da BASF assumem um comportamento positivo, sendo que as variações que ocorrem com mais frequência são positivas (ver figura 6.3 dos anexos).

Bayer

Com dados diários, ao longo do período em análise, a Bayer apresenta uma baixa similaridade com as restantes empresas, contudo o movimento dos preços de fecho das suas ações são mais semelhantes com a Siemens (0,59) e a Basf (0,58) e menos semelhantes com o USA Oil Fund (0,37) e com o S&P500 (0,47). Ao passar para variações semanais as empresas nas quais a Bayer apresenta maiores semelhanças são a Volkswagen (0,61), a BMW (0,57) e a Allianz (0,57), enquanto o S&P500 (0,32), a Lufthansa (0,35) e o USA Oil Fund (0,35) são as mais dissimilares.

Duas das sequências que ocorrem com um suporte superior a 80% nos movimentos dos preços das ações da Bayer são: $\langle E \rightarrow A \rightarrow A \rightarrow E \rangle$ e $\langle E \rightarrow D \rightarrow E \rightarrow A \rangle$ (ver tabela 6.7 dos anexos).

BMW

A análise de similaridade com as variações diárias indica que a empresa que apresenta maior similaridade com a BMW é a Daimler (0,61) e as entidades que apresentam mais dissemelhanças são o USA Oil Fund (0,43) e o S&P500 (0,47).

Com as variações semanais, o comportamento dos preços da BMW apresenta fortes semelhanças com duas outras empresas do mesmo setor, a Volkswagen (0,71) e a Continental (0,61). Por outro lado, apresenta pouca semelhança com o índice S&P500 (0,23), com o USA Oil Fund (0,38), e com as empresas Linde (0,38) e Lufthansa (0,38).

As sequências frequentes máximas encontradas nas variações semanais dos preços da BMW podem ser verificadas na tabela 6.7 dos anexos. Algumas dessas sequências são: $\langle C \rightarrow A \rightarrow A \rightarrow A \rangle$, $\langle D \rightarrow E \rightarrow A \rightarrow A \rangle$ e $\langle E \rightarrow A \rightarrow A \rightarrow E \rangle$, em que, por exemplo, a primeira indica que, com frequência superior a 80%, após a ocorrência de uma variação dos preços entre -2% e -1% e de dois aumentos dos preços consecutivos superiores a 2%, os preços voltam a aumentar mais de 2%, num intervalo temporal máximo de uma semana. Por outro lado, a frequência dos itens nas variações das sequências (ver figura 6.5 dos anexos) indica um comportamento positivo dos preços das ações da BMW durante o período em análise.

Commerzbank

A empresa mais semelhante ao Commerzbank, com as variações semanais, é a Siemens (0,61), por outro lado, as entidades S&P500 (0,29), a Fresenius (0,35), o USA Oil Fund (0,38), a Linde (0,38) e a Lufthansa (0,38) apresentam poucas semelhanças. Com variações diárias, a similaridade do Commerzbank com as outras empresas é pouco relevante, sendo igual ou inferior a 0,5. Contudo, destaca-se a dissemelhança com o USA Oil Fund (0,37), tal como nas variações semanais.

O algoritmo encontrou, com variações semanais, cerca de vinte cinco sequências frequentes máximas nos dados do Commerzbank (ver tabela 6.3 dos anexos). O comportamento dos preços da Commerzbank, durante o período em análise, evidenciam uma tendência positiva, sendo que as variações superiores a 2% (item “A”) ocorrem com uma frequência de 33%. No entanto, as variações dos preços inferiores a -2 % (item “D”) apresentam uma frequência de 24%, o que indica fortes variações nos preços durante o período em análise.

Continental

Tal como referido anteriormente, a análise de similaridade permite a identificação de alguns setores de atividade, nomeadamente do setor automóvel. Estas conclusões são evidentes pelo grau de similaridade, nos dados semanais, entre a Continental e a Volkswagen (0,61), a BMW (0,61) e a Daimler (0,57). Por outro lado apresenta uma reduzida similaridade com o índice S&P500 (0,29) e com a Linde (0,35). Com dados diários, a Volkswagen (0,55) e a Daimler (0,55) são as empresas que apresentam maior similaridade com a Continental. Note-se, contudo, que, com dados diários, o grau de similaridade destas duas empresas com a Continental é menor do que com dados semanais.

A tabela 6.4 dos anexos apresenta as sequências frequentes máximas encontradas nas variações diárias dos preços da Continental. À semelhança do ocorrido com a Volkswagen e de acordo com a forte similaridade das duas empresas, os preços de fecho da Continental apresentam um comportamento positivo com 44% das variações referentes a aumentos nos preços superiores a 2%. As variações dos preços inferiores a -2% (item “D”) ocorrem com uma frequência de 31% (figura 6.7 dos anexos).

Daimler

A Daimler é uma empresa que atua no setor de automóveis e componentes e como tal apresenta uma elevada similaridade com uma outra empresa do setor: a Volkswagen (0,76). No que diz respeito à dissemelhança, tal como o verificado nas empresas anteriormente analisadas, apresenta um comportamento pouco semelhante com o índice S&P500 (0,26) e com o USA Oil Fund (0,35). Com dados diários, a similaridade com a Volkswagen baixa para 0,59, registando-se outras similaridades elevadas, como é o caso da Allianz (0,61) e da BMW (0,61).

As sequências frequentes máximas encontradas, nos dados diários, com suporte mínimo de 80% podem ser verificadas na tabela 4 dos anexos. Algumas dessas sequências são: $\langle A \rightarrow A \rightarrow A \rightarrow E \rangle$, $\langle A \rightarrow B \rightarrow A \rightarrow B \rangle$ e $\langle A \rightarrow E \rightarrow D \rightarrow A \rangle$. Por exemplo, no caso da primeira, verifica-se que, com uma frequência mínima de 80%, após três ocorrências consecutivas de aumentos no preço superiores a 2%, ocorre, no intervalo máximo de uma semana, uma estabilização do preço entre -1% e 1%.

Deutsche Bank

As variações dos preços das ações do Deutsche Bank, no período de Setembro de 2014 a Março de 2015, não apresentam similaridades relevantes com as restantes empresas tanto com dados diários como com dados semanais, à excepção da BMW (0,58) no caso dos dados diários. Destaca-se, contudo, à semelhança do ocorrido anteriormente, uma maior dissemelhança com o índice S&P500 (0,32), com dados semanais.

O algoritmo encontrou cerca de vinte e seis sequências frequentes máximas com dados semanais (ver tabela 6.4 dos anexos). Algumas dessas sequências, que apresentam um suporte superior a 80% são: $\langle D \rightarrow D \rightarrow A \rightarrow E \rangle$, $\langle A \rightarrow E \rightarrow A \rightarrow D \rangle$ e $\langle B \rightarrow D \rightarrow C \rightarrow A \rangle$. No caso desta última, por exemplo, com frequência superior a 80%, ocorre um aumento do preço entre 1% e 2% (item “B”), seguindo-se uma variação inferior a -2% (item “D”), uma diminuição entre -2% e -1% (item “C”) e um aumento do preço superior a 2% (item “A”).

Lufthansa

As variações diárias e semanais dos preços de fecho da Lufthansa apresentam muito pouca semelhança com as restantes empresas em análise. Pelo contrário, estas apresentam grandes dissemelhanças com a mesma, destaca-se, por exemplo, com dados semanais, a reduzida semelhança com a Basf (0,29), com Fresenius (0,32), Linde (0,32), S&P500 (0,32), Bayer (0,35), Deutsche Telekom (0,35), Merck (0,35) e o USA Oil Fund (0,35).

A dissemelhança dos movimentos dos preços da Lufthansa com as restantes entidades é também indicada pela frequência das variações ocorridas durante o período em análise. Até aqui verificou-se que as empresas analisadas apresentavam, em geral, um comportamento positivo dos preços. Nos dados da Lufthansa verificam-se algumas diferenças. As variações que ocorrem com mais frequência são as inferiores a -2% (item “D”), seguindo-se variações entre 1% e 2% (item “B”) e os aumentos dos preços superiores a 2% (item “A”), com 34%, 28%, 27% de frequência, respetivamente (figura 6.10 dos anexos). Esta situação indica um comportamento mais oscilatório dos preços de fecho da Lufthansa, no período em análise. As sequências frequentes máximas da

Lufthansa encontradas com dados semanais podem ser visualizadas na tabela 6.7 dos anexos.

Deutsche Telekom

A empresa que mais semelhança apresenta com a Deutsche Telekom, relativamente às variações semanais de preços de fecho, é a Allianz (0,61). Por outro lado, as que mais se distinguem dos movimentos de preços da Deutsche Telekom são o índice S&P500 (0,29), a empresa Linde (0,32) e a Lufthansa (0,35). Com dados diários verifica-se algumas semelhanças com a Siemens (0,55), e com a BMW (0,55). A similaridade mais evidente ocorrem com o fundo de investimentos USA Oil Fund (0,38).

Foram encontradas vinte e uma sequências frequentes máximas nos movimentos de preços (tabela 6.4 dos anexos). Durante o período em análise, as variações dos preços de fecho da Deutsche Telekom assumem um comportamento positivo. Nas sequências encontradas os aumentos dos preços superiores a 2% (item “A”) ocorrem com uma frequência de 43%, seguindo-se as estabilizações dos preços entre -1% e 1% (item “E”) e as variações inferiores a -2% (item “D”), ambos com frequência de 21% (figura 6.11 dos anexos).

Fresenius SE&CO

As variações semanais dos preços das ações da Fresenius, durante o período em análise, apresentam similaridades muito baixas com as restantes empresas. Algumas das entidades onde ocorrem maiores dissemelhanças são no fundo de investimentos USA Oil Fund (0,32), na Lufthansa (0,32) e no Commerzbank (0,35). Com dados diários verificam-se similaridades relevantes com a Allianz (0,6) e a Siemens (0,6) e uma elevada dissemelhança com os movimentos dos preços do USA Oil Fund (0,36)

Tal como ocorrido nas empresas anteriormente analisada, também os movimentos de preços da Fresenius apresentam um comportamento crescente. Nas sequências frequentes máximas encontradas (tabela 6.5 dos anexos) verifica-se uma maior frequência de variações superiores a 2% (39% das ocorrências) seguindo-se os movimentos estáveis dos preços (item “E”, variações entre -1% e 1%), em 23% das ocorrências (figura 6.12 dos anexos). Algumas das sequências encontradas são: $\langle C \rightarrow D \rightarrow A \rightarrow E \rangle$, $\langle E \rightarrow E \rightarrow E \rightarrow C \rangle$ e $\langle B \rightarrow A \rightarrow B \rightarrow A \rangle$. Neste último caso, com um suporte mínimo de 80%,

após a ocorrência de um aumento do preço entre 1% e 2% (item “B”), seguindo-se, um aumento superior a 2% (item “A”) e, novamente uma variação entre 1% e 2% (item “B”), irá ocorrer um aumento superior a 2%, no intervalo máximo de uma semana (item “A”).

Heidelberg Cement

As principais semelhanças dos movimentos semanais dos preços de fecho da Heidelberg Cement ocorrem com a BMW (0,61), com a Continental (0,57) e com a Volkswagen (0,57), as três anteriores pertencentes ao setor automóvel. No que diz respeito às dissemelhanças, destaca-se as mais relevantes ocorridas com o índice S&P500 (0,32), e com a Fresenius (0,38). Com dados diários, surge a Daimler (0,58) e a Allianz (0,58) como as mais semelhantes e o fundo de investimentos USA Oil Fund (0,37) como o mais dissemelhante.

O algoritmo encontrou cerca de vinte e sete sequências frequentes máximas (tabela 6.5 dos anexos), entre as quais: $\langle E \rightarrow D \rightarrow A \rightarrow E \rangle$, $\langle D \rightarrow A \rightarrow E \rightarrow A \rangle$ e $\langle A \rightarrow A \rightarrow C \rightarrow A \rangle$. Durante o período em análise, verifica-se um comportamento positivo dos preços de fecho da Heidelberg Cement, com as variações superiores a 2% a ocorrerem com 41% de frequência, seguindo-se as variações inferiores a -2% com 25% de frequência (figura 6.13 dos anexos).

Linde

A Linde apresenta poucas semelhanças com os movimentos semanais de preços registados nas restantes entidades. A similaridade máxima que apresenta é de 0,45 e a mais baixa é de 0,32 com as empresas Daimler, Lufthansa e Deutsche Telekom. Contudo, com variações diárias, apresenta uma similaridade elevada com a Allianz (0,66) e com o S&P500 (0,63) e uma similaridade reduzida com o USA Oil Fund (0,34).

As sequências frequentes máximas encontradas nas variações de preços da Linde (tabela 6.5 dos anexos) indicam um comportamento dos preços positivo, durante o período em análise. No entanto, nas empresas anteriores verificou-se que as variações que ocorrem com mais frequência são aumentos do preço superiores a 2%, na Linde, as variações mais frequentes estão entre 1% e 2% (item “B”), com uma ocorrência de 27%, seguindo-se, com 26%, os aumentos superiores a 2% (item “A”) e com 22% as variações entre -1% e 1% (item “E”) (figura 6.14 dos anexos).

Merck

Com dados semanais, tal como verificado na Linde, também as variações semanais dos preços de fecho das ações da Merck apresentam pouca similaridade com as variações das restantes empresas. Destaca-se a dissemelhança com o USA Oil Fund (0,29), e com a empresa Lufthansa (0,35). Com dados diários, a Merck apresenta uma maior similaridade com a Linde (0,6), com a Siemens (0,58) e com a Basf (0,58) e menos similaridade com o USA Oil Fund (0,37).

Algumas das sequências frequentes máximas encontradas nos movimentos de preços da Merck (tabela 6.5 dos anexos) são: $\langle E \rightarrow D \rightarrow A \rightarrow E \rangle$, $\langle E \rightarrow E \rightarrow D \rightarrow C \rangle$ e $\langle E \rightarrow C \rightarrow A \rightarrow A \rangle$. Esta última sequência indica que, com um suporte mínimo de 80%, após uma variação do preço entre -1% e 1% (item “E”), seguindo-se uma variação entre -2% e -1% (item “C”) e um aumento superior a 2% (item “A”), segue-se um aumento do preço superior a 2% (item “A”), num intervalo máximo de uma semana. Os movimentos dos preços presentes nas sequências frequentes máximas indicam, no período em análise, um comportamento crescente com o item “A” (aumentos do preço superiores a 2%) a ocorrer em 44% das variações, sendo o item “E” (variações entre -1% e 1%) o segundo com mais ocorrências (20%), o “D” (variações inferiores a -2%) e o “B” (variações entre 1% e 2%) ambos com 13% e o C (variações entre -2% e -1%) com apenas 10% (figura 6.15 dos anexos).

Siemens

Os movimentos semanais de preços de fecho das ações da Siemens, durante o período de análise, apresentam similaridades relevantes com a Allianz (0,61) e o Commerzbank (0,61). Por outro lado, apresentam reduzida similaridade com o índice S&P500 (0,38) e com o USA Oil Fund (0,38). Com dados diários, verifica-se uma maior similaridade com as empresas Basf (0,68), Allianz (0,62), com a Fresenius (0,6) e a Linde (0,6). A maior dissemelhança ocorre com o USA Oil Fund (0,37).

O comportamento dos preços de fecho da Siemens nas sequências frequentes máximas, durante o período em análise, evidencia uma maior ocorrência de aumentos superiores a 2% (34% das variações), seguindo-se as variações entre -1% e 1%, que

ocorrem com 29% de frequência (figura 6.16 dos anexos). As sequências frequentes máximas podem ser visualizadas na tabela 6.6 dos anexos.

Volkswagen

Na Volkswagen, sendo uma empresa que atua no setor automóvel, as variações semanais dos preços de fecho das suas ações apresentam uma elevada similaridade com as variações semanais das restantes empresas do mesmo setor, tais como a BMW (0,71), a Continental (0,61) e a Daimler (0,76). Além destas, ainda apresenta uma similaridade relevante com a Bayer (0,61). Verifica-se uma baixa similaridade com a Linde (0,41) e, à semelhança do ocorrido com as restantes empresas, com o índice S&P500 (0,29) e com o USA Oil Fund (0,35). Com dados diários, verifica-se novamente uma similaridade relevante com Daimler (0,59), embora com menor grau que com dados semanais.

O algoritmo encontrou vinte e três sequências frequentes máximas nos movimentos dos preços da Volkswagen (tabela 6.6 dos anexos). Estas sequências indicam um comportamento positivo dos preços, durante o período em análise (figura 6.17 dos anexos). Os aumentos superiores a 2% (item “A”) são os que registam uma maior ocorrência (47% das variações). Algumas das sequências frequentes que caracterizam os movimentos dos preços da Volkswagen são: $\langle A \rightarrow E \rightarrow A \rightarrow B \rangle$, $\langle D \rightarrow D \rightarrow A \rightarrow E \rangle$ e $\langle E \rightarrow A \rightarrow A \rightarrow D \rangle$. Na primeira, por exemplo, com um suporte mínimo de 80%, após a ocorrência de um aumento do preço superior a 2% (item “A”), uma estabilização do preço entre -1% e 1% (item “E”) e, novamente, um aumento do preço superior a 2% (item “A”), ocorre um aumento entre 1% e 2% (item “B”), num intervalo temporal máximo de uma semana.

S&P500

Depois de analisadas as empresas que compõem o DAX, verificou-se que as variações semanais dos preços de fecho das mesmas apresentam uma similaridade muito reduzida com as variações semanais dos preços de fecho do índice S&P500, durante o período em análise. Neste sentido, conclui-se que o índice S&P500 apresenta poucas semelhanças com as empresas analisadas que compõe o DAX. Com dados semanais a similaridade do S&P500 com as restantes entidades é igual ou inferior a 0,41. Pelo

contrário, com dados diários, o S&P500 apresenta uma similaridade elevado com a Linde (0,63) e com a Allianz (0,62) e uma maior dissemelhança com o USA Oil Fund (0,31).

A reduzida similaridade com as restantes entidades, nas variações semanais, indicam que o comportamento dos preços de fecho do índice S&P500 não irão apresentar uma tendência positiva, tal como verificado nas restantes empresas, mas sim uma maior estabilidade, durante o período em análise (figura 6.18 dos anexos). As variações que mais ocorrem, nas sequências frequentes máximas encontradas, situam-se entre -1% e 1% (item “E”) o que indica uma maior estabilidade dos preços (47% das variações ocorridas).

As sequências frequentes máximas encontradas nos movimentos de preços do S&P500 podem ser visualizadas na tabela 6.6 dos anexos.

USA Oil Fund

Os movimentos dos preços de fecho do USA Oil Fund revelam, à semelhança do verificado com o índice S&P500, muito pouca similaridade com as empresas do DAX, tanto com dados diários como com dados semanais.

Verificou-se que as empresas negociadas no DAX apresentam, em geral, no período em análise, um comportamento positivo dos preços com grande parte das variações dos preços superiores a 2%, por sua vez, o índice S&P500 apresenta um comportamento mais estável, com um maior número de ocorrência entre -1% e 1%. As sequências frequentes máximas encontradas, com dados semanais, nas variações de preços do USA Oil Fund, registam uma tendência decrescente dos preços, com as variações inferiores a -2% (item “D”) a ocorrerem com 51% de frequência (figura 6.19 dos anexos). As sequências frequentes máximas encontradas foram registadas na tabela 6.6 dos anexos.

No estudo individualizado que foi realizado verificam-se alguns aspetos interessantes. Em primeiro lugar as ações negociadas no DAX apresentam, em geral, no período em análise, um comportamento positivo dos preços, com um maior número de ocorrências relativas a aumentos superiores a 2% nos preços semanais. Em segundo lugar, o índice S&P500 que apresenta maiores similaridades com as restantes entidades quando são usados dados diários em vez de dados semanais, evidencia um comportamento mais

estável, com uma maior ocorrência das variações entre -1% e 1% (com dados semanais). Em terceiro lugar, o fundo de investimentos USA Oil Fund apresenta uma reduzida similaridade com as restantes entidades, tanto com dados diários como com dados semanais, o que revela um comportamento decrescente dos preços, com uma maior ocorrência do item “D” (variações do preço inferiores a -2%).

Foi ainda possível verificar que os instrumentos financeiros europeus (empresas do DAX) e norte-americanos (S&P500 e USA Oil Fund) presentes na análise apresentam, em geral, muito poucas semelhanças. Tendo sido um dos objetivos desta dissertação, a tentativa de relacionar os movimentos de preços ocorridos entre estes dois grupos de entidades, conclui-se que os mesmos não evidenciam relações fortes entre eles. Contudo, e tal como foi explicado na secção 3.4.3.2, o que foi estudado nestas duas últimas secções foram os movimentos dos preços no mesmo instante temporal, ou seja, no mesmo dia/semana, o que não invalida que possam existir relações fortes nos movimentos de preços *à posteriori* entre as entidades.

Foi, ainda, possível identificar semelhanças importantes entre empresas no mesmo setor atividade, como no setor automóvel, as quais eram marcadas com níveis de similaridade elevados e, consequentemente, por tendências idênticas nos movimentos de preços.

Com a análise realizada nesta secção é possível obter um conjunto de informação mais detalhada sobre o comportamento de cada entidade presente neste estudo, durante o período em análise.

4 Conclusões e Trabalho Futuro

Neste capítulo serão apresentadas as principais conclusões da dissertação bem como sugestões para trabalhos futuros.

4.1 Conclusões

Este trabalho teve como objetivo a análise das variações de preços de um conjunto de entidades negociadas nos mercados financeiros, nomeadamente 17 empresas presentes no DAX, o índice S&P500 e o fundo de investimento USA Oil Fund, durante o período compreendido entre Setembro de 2014 e Setembro de 2015. Para tal foi utilizada a técnica *Sequence Pattern Mining* com o objetivo de extrair padrões sequenciais que caracterizassem os movimentos dos preços das entidades referidas anteriormente, quer ao nível coletivo (várias entidades), quer ao nível individual (uma entidade). O objetivo do estudo era fornecer informação útil de forma a apoiar o processo de decisão dos investidores. Adicionalmente, foi realizada a análise do impacto da granularidade temporal dos dados, a granularidade de discretização das variações dos preços e a aprendizagem de sequências com restrições nos padrões obtidos.

Foi possível a obtenção de padrões sequenciais que descrevem o comportamento dos preços das 19 entidades em análise, com um bom desempenho do algoritmo (tempos de execução baixos). Esta informação, que caracteriza os movimentos dos preços durante sete meses, produz conhecimento útil para o investidor, podendo ser, no futuro, usada para decisões de investimento.

O uso de restrições no processo de aprendizagem de sequências é um procedimento essencial tanto no desempenho do algoritmo como na “condução” da análise para o objetivo do estudo. Em primeiro lugar, verificou-se que a definição de um suporte mínimo e do número máximo de itens presentes nas sequências produzia uma grande quantidade de sequências frequentes, tornando intratável o processo de descoberta de conhecimento. Optou-se, assim, pela implementação de restrições ao nível da

diferença temporal máxima de ocorrência dos itens numa sequência o que permitiu, por um lado, a melhoria do desempenho do algoritmo, através da redução do espaço de procura e, consequentemente, do tempo de processamento, por outro lado, conduziu à obtenção de sequências que melhor caracterizam os preços das ações, e, assim, à produção de conhecimento útil para o investidor.

O estudo da granularidade da discretização das variações dos preços teve como objetivo a análise da estabilidade/instabilidade dos movimentos dos preços das entidades. O uso de cinco categorias de variação dos preços permitiu compreender melhor o comportamento destas entidades do que com apenas três categorias. Verificou-se que, durante os sete meses em análise, em geral, os preços indicam uma tendência positiva, com grande parte das variações ocorridas com aumentos acima dos 2%. Contudo, importa referir que esta tendência é essencialmente causada pelas empresas negociadas no DAX, uma vez que, no estudo individualizado das entidades concluiu-se que o índice S&P500 apresenta um comportamento relativamente constante e o fundo de investimentos USA Oil Fund revela uma tendência negativa, durante o período em análise.

A granularidade temporal dos dados é um dos aspetos que mais altera os resultados obtidos. O facto de os dados diários incluírem um número maior de transações que os dados semanais conduz a um espaço de procura maior e, consequentemente, aumenta o tempo de processamento do algoritmo. Visto que os padrões sequenciais encontrados apresentam variações consecutivas dos preços ocorridas no máximo em 1 dia/semana e, consequentemente, os padrões sequenciais com dados diários indicam uma tomada de decisão mais rápida por parte do investidor, os dados semanais podem representar, na prática, uma melhor opção para os investidores.

Na análise de similaridade, o efeito da granularidade temporal dos dados produz efeitos diferentes. Com dados diários, as semelhanças/dissemelhanças no comportamento dos preços são menos evidentes do que com dados semanais como, por exemplo, a similaridade das empresas do setor automóvel é mais notória com dados semanais do que com dados diários. Note-se que, a análise de similaridade tem um papel muito importante num dos objetivos desta dissertação: fornecer informação útil sobre o comportamento das ações nos mercados financeiros. Quando duas entidades apresentam um grau elevado de similaridade indica que os movimentos dos preços de fecho das ações dessas mesmas

empresas exibem um comportamento muito semelhante. O investidor pode tirar partido desta informação no seu processo de tomada de decisão de investimentos, visto que, segundo a teoria económica, a fim de diminuir o risco da carteira é fundamentalmente diversificar a mesma, o que passa pela aposta em ações que apresentam comportamentos dissemelhantes.

De referir, ainda, que a inclusão do índice S&P500 e do USA Oil Fund na análise teve como objetivo a tentativa de descobrir relações entre os instrumentos financeiros norte-americanos e europeus. Verificou-se que não é possível encontrar relações entre estes instrumentos, tanto os preços de fecho do S&P500 como do USA Oil Fund apresentam grandes dissemelhanças com as restantes ações negociadas no DAX. Contudo, importa esclarecer que, o facto de os movimentos de preços de algumas empresas apresentarem reduzidas similaridades não invalida que possam existir relações fortes nos movimentos de preços *à posteriori* entre as entidades, uma vez que a medida de similaridade usada compara os movimentos dos preços ocorridos no mesmo instante temporal.

O desenvolvimento deste trabalho contribuiu positivamente para a produção de conhecimento que pode ser usado na tomada de decisão dos investidores. Um investidor que opte pela constituição de uma carteira de investimentos composta por algumas das empresas/entidades presentes nesta análise obtém um conjunto de informação relevante no que diz respeito às variações dos preços das ações e à relação entre as mesmas. Por outro lado, esta análise pode ser aplicada a outras entidades que sejam da preferência dos investidores. A eficiência deste estudo fica patente com um tempo de processamento necessário para a descoberta de padrões sequenciais relativamente baixo tendo em conta a quantidade de dados em análise. No caso das variações semanais o algoritmo demora cerca de 0,27 segundos a encontrar as sequências frequentes (com suporte 80% e com as restrições adicionais) e com dados diários o tempo de processamento aumenta ligeiramente para 2,75, embora ainda bastante aceitável. Por outro lado, este trabalho pode representar um importante ponto de partida para outros na mesma área.

O *SPADE* é um dos algoritmos, baseados na metodologia *Apriori*, que melhor desempenho apresenta. A justificação para tal deve-se ao facto do seu desenvolvimento

ter ocorrido com a finalidade de ultrapassar algumas das limitações dos algoritmos baseados no *Apriori*, nomeadamente no que se refere ao custo das consecutivas passagens pela base de dados. O *SPADE*, ao contrário dos restantes algoritmos *Apriori*, usa propriedades combinatórias para decompor o problema original em subproblemas menores, que podem ser independentemente resolvidos na memória principal usando técnicas de pesquisa *lattice*, e usando operações de agregação simples. Contudo, face ao problema da geração de um grande número de sequências frequentes, optou-se pela escolha do *cSPADE* (*constrained SPADE*) que permite a incorporação de restrições adicionais a fim de reduzir o número de sequências frequentes encontradas.

As principais limitações encontradas na realização desta dissertação foram a escolha do período de análise compreendido entre Setembro de 2014 e Março de 2015 e que corresponde apenas a sete meses de dados. Poderia ter sido interessante, do ponto de vista do estudo da granularidade temporal, analisar um período mais longo (talvez um ano). Adicionalmente, a escolha da discretização das variações pode ter limitado os resultados obtidos, nomeadamente na inclusão da identificação das variações dos preços ocorridas entre -1% e 1% que podem não representar informação relevante. Por outro lado, a limitação temporal para a realização da dissertação revelou-se um fator restritivo para a realização de outras experiências que poderiam ter sido importantes para o presente estudo.

4.2 Trabalho Futuro

Este trabalho produz um conjunto de informação útil do ponto de visto do investidor, podendo ser usada, posteriormente, para previsão dos movimentos dos preços das ações. A informação descoberta pode ser utilizada para a construção de um sistema de suporte à decisão que indique ao investidor a compra ou venda de determinada ação com base na informação produzida.

Por outro lado, pode ser interessante a análise de outros instrumentos financeiros diferentes dos incluídos neste estudo, consoante as preferências dos investidores. Adicionalmente, poderá ser de explorar a utilização de outros algoritmos, um período

temporal diferente bem como outro tipo de discretização das variações dos preços que não incluía as variações entre -1% e 1% e que analise as variações superiores a 2% e inferiores a -2%.

5 Referências Bibliográficas

- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *ACM SIGMOD Record* (Vol. 22, No. 2, pp. 207-216). ACM.
- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).
- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on* (pp. 3-14). IEEE.
- Ayres, J., Flannick, J., Gehrke, J., & Yiu, T. (2002). Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 429-435). ACM.
- Antunes, C., & Oliveira, A. L. (2004, March). Sequential pattern mining with approximated constraints. In *Int. Conf Applied Computing* (pp. 131-138).
- Chakrabarti, S., Sarawagi, S., & Dom, B. (1998). Mining surprising patterns using temporal description length. In *VLDB* (Vol. 98, pp. 606-617).
- Das, G., Lin, K. I., Mannila, H., Renganathan, G., & Smyth, P. (1998). Rule Discovery from Time Series. In *KDD* (Vol. 98, pp. 16-22).
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work*. *The journal of Finance*, 25(2), (pp. 383-417).
- Kalaivany, M., & Uma, V. (2014) Mining Sequential Patterns for interval based events by applying multiple constraints. *International Journal on Computational Sciences & Applications (IJCSA)* (Vol. 4, pp. 59-66).

Leigh, W., Modani, N., Purvis, R., & Roberts, T. (2002). Stock market trading rule discovery using technical charting heuristics. *Expert Systems with Applications*, 23(2), (pp. 155-159).

Leleu, M., & Boulicaut, J. (2002). Signing stock market situations by means of characteristic sequential patterns. In *Proc. of the 3rd International Conference on Data Mining (DM'02)*, (pp. 655-664).

Liu, J. N., & Kwong, R. W. (2007). Automatic extraction and identification of chart patterns towards financial forecast. *Applied Soft Computing*, 7(4), (pp. 1197-1208).

Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. In *Proc. 2000 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA. ACM Press. (pp. 1-12).

Marketos, G. D., Pediaditakis, K., Theodoridis, Y., & Theodoulidis, B. (2004). *Intelligent stock market assistant using temporal data mining*. University of Manchester.

Markowitz, H. (1952). *Portfolio selection**. *The journal of finance*, 7(1), (pp. 77-91)

Martínez-Álvarez, F., Troncoso, A., Riquelme, J. C., & Aguilar-Ruiz, J. S. (2008, December). LBF: A labeled-based forecasting algorithm and its application to electricity price time series. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on* (pp. 453-461). IEEE.

Mitsa, T. (2010). *Temporal data mining*. CRC Press.

Mooney, C. H., & Roddick, J. F. (2013). Sequential pattern mining--approaches and algorithms. *ACM Computing Surveys (CSUR)*, 45(2), (pp 1-46).

- Parracho, P., Neves, R., & Horta, N. (2010). Trading in financial markets using pattern recognition optimized by genetic algorithms. In *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation* (pp. 2105-2106). ACM.
- Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. C. (2001a). Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *icccn* (p. 0215). IEEE.
- Pei, J., Han, J., & Lakshmanan, L. V. (2001b). Mining frequent itemsets with convertible constraints. In *Data Engineering, 2001. Proceedings. 17th International Conference on* (pp. 433-442). IEEE.
- Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q. & Hsu, M. C. (2004). Mining sequential patterns by pattern-growth: The prefixspan approach. *Knowledge and Data Engineering, IEEE Transactions on*, 16 (11), (pp. 1424-1440).
- Povinelli, R. J. (2001). Identifying temporal patterns for characterization and prediction of financial time series events. In *Temporal, Spatial, and Spatio-Temporal Data Mining* Springer Berlin Heidelberg. (pp. 46-61).
- Slimani, T., & Lazzez, A. (2013). Sequential mining: patterns and algorithms analysis. *arXiv preprint arXiv:1311.0350*.
- Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. Springer Berlin Heidelberg. (pp. 1-17).
- Vishwanath, R. H., Leena, S., Srikantaiah, K. C., Kumar, K. S., Shenoy, P. D., Venugopal, K. R.; & Patnaik, L. M. (2013). Forecasting Stock Time-Series using Data Approximation and Pattern Sequence Similarity. *arXiv preprint arXiv:1309.2517*.
- Wang, W., & Yang, J. (2006). Mining sequential patterns from large data sets (Vol. 28). Springer Science & Business Media.

Wang, J. L., & Chan, S. H. (2009). Trading rule discovery in the US stock market: An empirical study. *Expert Systems with Applications*, 36(3), (pp. 5450-5455).

Wu, K. P., Wu, Y. P., & Lee, H. M. (2014). Stock Trend Prediction by Using K-Means and AprioriAll Algorithm for Sequential Chart Pattern Mining. *Journal of Information Science and Engineering*, 30, (pp. 653-667).

Zaki, M. J. (2000). Sequence mining in categorical domains: incorporating constraints. In *Proceedings of the ninth international conference on Information and knowledge management*. ACM. (pp. 422-429).

Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine learning*, 42(1-2), (pp. 31-60).

Zhang, Z., Jiang, J., Liu, X., Lau, R., Wang, H., & Zhang, R. (2010). A real time hybrid pattern matching scheme for stock time series. In *Proceedings of the Twenty-First Australasian Conference on Database Technologies-Volume 104*. Australian Computer Society, Inc. (pp. 161-170).

Yan, X., Han, J., & Afshar, R. (2003). CloSpan: Mining closed sequential patterns in large datasets. In *Proceedings of SIAM International Conference on Data Mining* (pp. 166-177).

6 Anexos

Código do R:

```
>library(Matrix)

>library(arules)

>library(arulesSequences)

>c<-read.delim("diariosC2.csv",header=FALSE)

>x<-read_baskets("diariosC2.csv", sep=";", info= c("sequenceID", "eventID", "SIZE"))

>d<-as(x, "data.frame")

###

>s2 <- cspade(x, parameter = list(support = 0.8,maxsize=7,maxlen=7), control = list(verbose =
TRUE,tidLists=TRUE))

>summary(s2)

####MINERAÇÃO DE SEQUÊNCIAS COM RESTRIÇÕES

>s4 <- cspade(x, parameter = list(support = 0.8,maxsize=7,maxlen=7,maxwin=4,maxgap=1),
control = list(verbose = TRUE,tidLists=TRUE))

>summary(s4)

>as(s4,"data.frame")

####MINERAÇÃO DE SEQUÊNCIAS MÁXIMAS

>p<-subset(s4,is.maximal(s4))

>i<-as(p,"data.frame")

#### ESTUDO DA SIMILARIDADE

>z <- as(x, "timedsequences")

>similarity(z)
```


Tabela 6.1: Sequências frequentes máximas com dados diários

[illegible]

[illegible]

[illegible]

[illegible]

<C→B→E→E→E→E→C>	0,98	<C→C→E→E→E→C→E>	0,89	<E→D→A→E→E→E→D>	0,84	<D→B→E→E→E→A→E>	0,79
<E→B→E→E→E→E→C>	0,98	<E→C→E→A→A→B→E>	0,89	<A→E→A→E→E→E→D>	0,84	<B→E→E→E→A→E→D>	0,79
<C→C→E→E→E→E→C>	0,98	<E→E→E→D→A→B→E>	0,89	<B→E→A→E→E→E→D>	0,84	<E→E→A→E→B→E→D>	0,79
<A→E→E→E→E→E→C>	0,98	<E→B→E→E→A→B→E>	0,89	<D→E→A→E→E→E→D>	0,84	<B→E→E→B→C→E→D>	0,79
<B→E→E→E→E→E→C>	0,98	<E→C→E→E→A→B→E>	0,89	<E→A→B→E→E→E→D>	0,84	<A→E→A→E→C→E→D>	0,79
<C→E→E→E→E→E→C>	0,98	<A→E→E→E→A→B→E>	0,89	<D→E→B→E→E→E→D>	0,84	<A→E→B→E→C→E→D>	0,79
<E→E→E→E→E→E→C>	0,98	<B→E→E→E→A→B→E>	0,89	<E→A→C→E→E→E→D>	0,84	<E→E→C→E→D→E→D>	0,79
<E→E→C→E→E→D→C>	0,98	<C→E→E→E→A→B→E>	0,89	<E→C→C→E→E→E→D>	0,84	<E→C→E→E→D→E→D>	0,79
<E→E→E→E→E→D→C>	0,98	<D→E→E→E→A→B→E>	0,89	<D→E→C→E→E→E→D>	0,84	<B→E→E→E→D→E→D>	0,79
<E→E→C→E→E→C→C>	0,98	<E→E→E→A→B→B→E>	0,89	<E→D→D→E→E→E→D>	0,84	<C→E→E→E→D→E→D>	0,79
<E→B→E→E→E→E→C>	0,98	<D→E→E→B→B→B→E>	0,89	<C→E→D→E→E→E→D>	0,84	<E→E→C→A→E→E→D>	0,79
<B→E→E→E→E→E→C>	0,98	<E→E→A→C→B→B→E>	0,89	<E→E→D→E→E→E→D>	0,84	<E→A→A→B→E→E→D>	0,79
<E→E→E→E→E→C→C>	0,98	<C→E→E→C→B→B→E>	0,89	<D→A→E→E→E→E→D>	0,84	<A→E→A→B→E→E→D>	0,79
<E→E→B→E→E→B→C>	0,98	<B→E→B→E→B→B→E>	0,89	<B→B→E→E→E→E→D>	0,84	<E→E→A→B→E→E→D>	0,79
<E→E→C→E→E→B→C>	0,98	<E→E→C→E→B→B→E>	0,89	<D→C→E→E→E→E→D>	0,84	<E→A→C→B→E→E→D>	0,79
<E→B→E→E→E→B→C>	0,98	<B→B→E→E→B→B→E>	0,89	<D→D→E→E→E→E→D>	0,84	<A→E→C→B→E→E→D>	0,79
<E→C→E→E→E→B→C>	0,98	<A→C→E→E→B→B→E>	0,89	<D→E→E→E→E→E→D>	0,84	<E→D→E→B→E→E→D>	0,79
<B→E→E→E→E→B→C>	0,98	<E→D→E→E→B→B→E>	0,89	<E→E→E→E→C→D→D>	0,84	<E→E→A→C→E→E→D>	0,79
<C→E→E→E→E→B→C>	0,98	<E→E→D→A→C→B→E>	0,89	<E→C→E→C→E→D→D>	0,84	<A→B→E→C→E→E→D>	0,79
<E→E→E→E→E→B→C>	0,98	<E→C→E→A→C→B→E>	0,89	<E→E→E→C→E→D→D>	0,84	<E→C→E→D→E→E→D>	0,79
<E→E→E→B→B→E→B>	0,98	<B→E→E→A→C→B→E>	0,89	<C→E→C→E→E→D→D>	0,84	<B→E→E→D→E→E→D>	0,79
<E→E→B→E→B→E→B>	0,98	<C→E→E→A→C→B→E>	0,89	<E→C→E→E→E→D→D>	0,84	<C→E→E→D→E→E→D>	0,79
<B→B→E→E→B→E→B>	0,98	<E→C→E→D→C→B→E>	0,89	<B→E→E→E→E→D→D>	0,84	<D→A→A→E→E→E→D>	0,79
<E→B→E→E→B→E→B>	0,98	<C→E→E→D→C→B→E>	0,89	<E→E→E→E→E→D→D>	0,84	<D→D→A→E→E→E→D>	0,79
<B→E→E→E→B→E→B>	0,98	<E→E→A→E→C→B→E>	0,89	<C→E→E→E→C→C→D>	0,84	<C→E→A→E→E→E→D>	0,79
<D→E→E→E→B→E→B>	0,98	<E→E→B→E→C→B→E>	0,89	<A→E→B→E→E→C→D>	0,84	<A→C→B→E→E→E→D>	0,79
<E→E→E→E→B→E→B>	0,98	<E→E→C→E→C→B→E>	0,89	<E→E→D→E→E→C→D>	0,84	<E→D→B→E→E→E→D>	0,79
<E→E→E→C→C→E→B>	0,98	<C→B→E→E→C→B→E>	0,89	<E→C→E→E→E→C→D>	0,84	<B→E→B→E→E→E→D>	0,79
<E→B→E→E→C→E→B>	0,98	<B→E→E→E→D→B→E>	0,89	<D→E→E→E→E→C→D>	0,84	<E→B→C→E→E→E→D>	0,79
<B→E→E→E→C→E→B>	0,98	<C→E→E→E→D→B→E>	0,89	<E→E→E→C→E→B→D>	0,84	<A→E→C→E→E→E→D>	0,79
<C→E→E→E→C→E→B>	0,98	<E→E→A→A→E→B→E>	0,89	<E→E→B→A→A→E→C>	0,84	<C→E→C→E→E→E→D>	0,79
<E→E→E→E→C→E→B>	0,98	<E→A→C→A→E→B→E>	0,89	<E→E→E→D→A→E→C>	0,84	<E→C→D→E→E→E→D>	0,79
<E→E→B→B→E→E→B>	0,98	<E→C→E→A→E→B→E>	0,89	<E→B→A→E→A→E→C>	0,84	<A→B→E→E→E→E→D>	0,79
<E→B→E→B→E→E→B>	0,98	<C→E→E→A→E→B→E>	0,89	<B→E→B→E→A→E→C>	0,84	<C→C→E→E→E→E→D>	0,79
<E→C→E→B→E→E→B>	0,98	<E→E→A→B→E→B→E>	0,89	<E→D→E→E→A→E→C>	0,84	<E→E→B→E→C→D→D>	0,79
<C→E→E→B→E→E→B>	0,98	<E→E→C→B→E→B→E>	0,89	<C→E→E→E→A→E→C>	0,84	<E→E→C→E→C→D→D>	0,79
<D→E→E→B→E→E→B>	0,98	<C→B→E→B→E→B→E>	0,89	<C→E→E→A→B→E→C>	0,84	<E→E→E→B→E→D→D>	0,79
<E→E→E→B→E→E→B>	0,98	<E→D→A→C→E→B→E>	0,89	<E→E→A→B→B→E→C>	0,84	<E→B→E→C→E→D→D>	0,79
<E→E→B→C→E→E→B>	0,98	<B→E→A→C→E→B→E>	0,89	<E→A→E→B→B→E→C>	0,84	<C→E→E→C→E→D→D>	0,79
<E→E→C→C→E→E→B>	0,98	<C→E→A→C→E→B→E>	0,89	<E→B→E→B→B→E→C>	0,84	<B→E→C→E→E→D→D>	0,79
<B→B→E→C→E→E→B>	0,98	<E→E→D→C→E→B→E>	0,89	<A→E→B→E→B→E→C>	0,84	<E→A→E→E→E→D→D>	0,79
<E→B→E→C→E→E→B>	0,98	<E→C→E→C→E→B→E>	0,89	<E→A→C→E→B→E→C>	0,84	<C→C→E→E→E→D→D>	0,79
<B→E→E→C→E→E→B>	0,98	<E→D→E→C→E→B→E>	0,89	<A→E→C→E→B→E→C>	0,84	<E→D→E→E→E→D→D>	0,79
<E→E→E→C→E→E→B>	0,98	<A→E→E→C→E→B→E>	0,89	<B→E→C→E→B→E→C>	0,84	<A→E→E→E→E→D→D>	0,79

[illegible]

<B→B→E→E→E→B→B>	0,98	<E→B→E→C→E→A→E>	0,89	<B→E→D→E→E→E→C>	0,84	<C→A→E→E→B→E→C>	0,79
<E→B→E→E→E→B→B>	0,98	<E→C→E→C→E→A→E>	0,89	<D→C→E→E→E→E→C>	0,84	<A→B→E→E→B→B→E→C>	0,79
<E→C→E→E→E→B→B>	0,98	<A→E→E→C→E→A→E>	0,89	<C→E→C→E→E→D→C>	0,84	<C→C→E→E→B→E→C>	0,79
<A→E→E→E→E→B→B>	0,98	<B→E→E→C→E→A→E>	0,89	<B→E→E→E→A→C→C>	0,84	<C→D→E→E→B→E→C>	0,79
<B→E→E→E→E→B→B>	0,98	<E→C→E→D→E→A→E>	0,89	<E→E→E→E→A→C→C>	0,84	<B→E→E→A→C→E→C>	0,79
<C→E→E→E→E→B→B>	0,98	<B→E→E→D→E→A→E>	0,89	<E→C→E→E→B→C→C>	0,84	<C→E→E→A→C→E→C>	0,79
<E→E→E→E→E→B→B>	0,98	<E→E→E→D→E→A→E>	0,89	<B→E→E→E→B→C→C>	0,84	<E→B→E→C→C→E→C>	0,79
<E→E→E→C→C→E→A>	0,98	<E→B→A→E→E→A→E>	0,89	<E→E→E→E→B→C→C>	0,84	<A→E→E→C→C→E→C>	0,79
<E→E→C→E→C→E→A>	0,98	<B→E→A→E→E→A→E>	0,89	<B→E→E→B→E→C→C>	0,84	<C→E→E→C→C→E→C>	0,79
<E→E→E→E→C→E→A>	0,98	<E→A→B→E→E→A→E>	0,89	<E→E→E→B→E→C→C>	0,84	<E→E→A→E→C→E→C>	0,79
<E→E→E→E→E→E→A>	0,98	<A→E→B→E→E→A→E>	0,89	<E→E→C→C→E→C→C>	0,84	<E→A→B→E→C→E→C>	0,79
<E→E→E→C→E→C→A>	0,98	<E→B→C→E→E→A→E>	0,89	<E→C→E→C→E→C→C>	0,84	<C→A→E→E→C→E→C>	0,79
<E→E→E→E→E→C→A>	0,98	<E→C→C→E→E→A→E>	0,89	<E→B→B→E→E→C→C>	0,84	<B→C→E→E→C→E→C>	0,79
<E→E→D→A→A→E→E>	0,95	<B→E→C→E→E→A→E>	0,89	<E→C→C→E→E→C→C>	0,84	<A→E→E→C→D→E→C>	0,79
<E→E→E→A→A→E→E>	0,95	<C→E→C→E→E→A→E>	0,89	<B→B→E→E→E→C→C>	0,84	<E→B→E→E→D→E→C>	0,79
<E→E→E→B→A→E→E>	0,95	<E→E→D→E→E→A→E>	0,89	<C→B→E→E→E→C→C>	0,84	<C→E→A→A→E→E→C>	0,79
<E→C→E→C→A→E→E>	0,95	<A→A→E→E→E→A→E>	0,89	<E→E→A→E→B→B→C>	0,84	<E→D→B→A→E→E→C>	0,79
<C→E→E→D→A→E→E>	0,95	<E→A→E→E→E→A→E>	0,89	<E→A→E→E→B→B→C>	0,84	<E→E→C→A→E→E→C>	0,79
<E→E→E→D→A→E→E>	0,95	<A→B→E→E→E→A→E>	0,89	<E→B→E→E→B→B→C>	0,84	<E→C→D→A→E→E→C>	0,79
<E→E→A→E→A→E→E>	0,95	<C→B→E→E→E→A→E>	0,89	<C→E→E→E→B→B→C>	0,84	<E→D→D→A→E→E→C>	0,79
<B→E→B→E→A→E→E>	0,95	<B→C→E→E→E→A→E>	0,89	<B→E→E→A→E→B→C>	0,84	<C→C→E→A→E→E→C>	0,79
<E→E→B→E→A→E→E>	0,95	<C→E→E→E→E→A→E>	0,89	<C→E→E→A→E→B→C>	0,84	<E→D→E→A→E→E→C>	0,79
<E→C→C→E→A→E→E>	0,95	<E→E→E→E→A→E→D>	0,89	<E→B→E→B→E→B→C>	0,84	<D→E→E→A→E→E→C>	0,79
<B→E→C→E→A→E→E>	0,95	<E→E→E→C→B→E→D>	0,89	<E→B→E→C→E→B→C>	0,84	<E→B→B→B→E→E→C>	0,79
<E→A→E→E→A→E→E>	0,95	<E→E→C→E→B→E→D>	0,89	<B→E→E→C→E→B→C>	0,84	<A→E→C→B→E→E→C>	0,79
<E→B→E→E→A→E→E>	0,95	<E→B→E→E→B→E→D>	0,89	<E→A→E→E→E→B→C>	0,84	<B→E→C→B→E→E→C>	0,79
<C→C→E→E→A→E→E>	0,95	<B→E→E→E→B→E→D>	0,89	<C→B→E→E→E→B→C>	0,84	<E→E→D→B→E→E→C>	0,79
<E→C→E→E→A→E→E>	0,95	<E→E→E→C→C→E→D>	0,89	<A→C→E→E→E→B→C>	0,84	<C→A→E→B→E→E→C>	0,79
<A→E→E→E→A→E→E>	0,95	<E→B→E→E→C→E→D>	0,89	<B→C→E→E→E→B→C>	0,84	<A→B→E→B→E→E→C>	0,79
<B→E→E→E→A→E→E>	0,95	<A→E→E→E→C→E→D>	0,89	<E→A→E→E→B→A→C>	0,84	<C→B→E→B→E→E→C>	0,79
<C→E→E→E→A→E→E>	0,95	<B→E→E→E→C→E→D>	0,89	<D→E→E→E→B→A→C>	0,84	<B→C→E→B→E→E→C>	0,79
<E→E→E→A→B→E→E>	0,95	<C→E→E→E→C→E→D>	0,89	<E→E→E→E→C→A→C>	0,84	<E→D→E→B→E→E→C>	0,79
<E→E→C→B→B→E→E>	0,95	<E→E→E→A→E→E→D>	0,89	<A→E→E→E→D→A→C>	0,84	<A→E→A→C→E→E→C>	0,79

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

<C→E→B→E→E→B→B>	0,95	<E→D→E→E→D→E→E>	0,84	<B→C→E→D→E→E→E>	0,79	<D→E→E→B→E→B→A>	0,79
<C→B→E→E→E→E→B→B>	0,95	<D→E→E→E→D→E→E>	0,84	<E→D→E→D→E→E→E>	0,79	<E→C→C→E→B→B→A>	0,79
<D→E→E→E→E→E→B→B>	0,95	<C→E→A→A→E→E→E>	0,84	<D→E→E→D→E→E→E>	0,79	<B→E→E→C→E→B→A>	0,79
<A→E→E→E→E→E→A→B>	0,95	<A→E→B→A→E→E→E>	0,84	<C→A→A→E→E→E→E>	0,79	<C→E→E→C→E→B→A>	0,79
<D→E→E→E→E→E→A→B>	0,95	<B→E→C→A→E→E→E>	0,84	<D→B→A→E→E→E→E>	0,79	<D→E→E→C→E→B→A>	0,79
<E→E→E→E→A→E→A>	0,95	<C→E→C→A→E→E→E>	0,84	<C→A→B→E→E→E→E>	0,79	<E→E→E→D→E→B→A>	0,79
<E→E→B→E→B→E→A>	0,95	<D→E→C→A→E→E→E>	0,84	<C→A→C→E→E→E→E>	0,79	<E→B→B→E→E→B→A>	0,79
<E→A→E→E→B→E→A>	0,95	<C→D→D→A→E→E→E>	0,84	<A→B→C→E→E→E→E>	0,79	<D→E→B→E→E→B→A>	0,79
<E→B→E→E→B→E→A>	0,95	<E→D→D→A→E→E→E>	0,84	<C→C→C→E→E→E→E>	0,79	<E→A→C→E→E→B→A>	0,79
<E→C→E→E→B→E→A>	0,95	<B→E→D→A→E→E→E>	0,84	<D→D→C→E→E→E→E>	0,79	<D→E→C→E→E→B→A>	0,79
<A→E→E→E→B→E→A>	0,95	<A→A→E→A→E→E→E>	0,84	<C→B→D→E→E→E→E>	0,79	<E→C→D→E→E→B→A>	0,79
<C→E→E→E→B→E→A>	0,95	<D→A→E→A→E→E→E>	0,84	<E→E→B→E→B→D→E>	0,79	<C→E→D→E→E→B→A>	0,79
<D→E→E→E→B→E→A>	0,95	<B→B→E→A→E→E→E>	0,84	<E→E→E→B→C→D→E>	0,79	<A→A→E→E→E→B→A>	0,79
<E→E→E→E→B→E→A>	0,95	<C→B→E→A→E→E→E>	0,84	<E→D→E→E→C→D→E>	0,79	<D→A→E→E→E→B→A>	0,79
<E→E→B→E→C→E→A>	0,95	<B→C→E→A→E→E→E>	0,84	<E→E→B→E→D→D→E>	0,79	<B→B→E→E→E→B→A>	0,79
<A→A→E→E→C→E→A>	0,95	<C→D→E→A→E→E→E>	0,84	<B→E→C→E→D→D→E>	0,79	<D→B→E→E→E→B→A>	0,79
<B→E→E→E→C→E→A>	0,95	<D→D→E→A→E→E→E>	0,84	<A→E→E→E→D→D→E>	0,79	<A→C→E→E→E→B→A>	0,79
<E→E→E→A→E→E→A>	0,95	<E→A→A→B→E→E→E>	0,84	<E→A→E→B→E→D→E>	0,79	<C→D→E→E→E→B→A>	0,79
<E→B→E→B→E→E→A>	0,95	<E→B→A→B→E→E→E>	0,84	<E→B→E→B→E→D→E>	0,79	<B→E→E→E→B→A→A>	0,79
<B→E→E→B→E→E→A>	0,95	<D→D→A→B→E→E→E>	0,84	<E→A→E→C→E→D→E>	0,79	<E→E→E→C→A→A>	0,79
<C→E→E→B→E→E→A>	0,95	<A→E→A→B→E→E→E>	0,84	<E→C→E→D→E→D→E>	0,79	<E→E→E→C→D→A→A>	0,79
<D→E→E→B→E→E→A>	0,95	<B→E→A→B→E→E→E>	0,84	<C→E→E→D→E→D→E>	0,79	<B→E→E→D→D→A→A>	0,79
<E→E→E→B→E→E→A>	0,95	<D→A→C→B→E→E→E>	0,84	<E→C→A→E→E→D→E>	0,79	<C→E→E→D→D→A→A>	0,79
<E→E→E→C→E→E→A>	0,95	<B→E→D→B→E→E→E>	0,84	<E→A→B→E→E→D→E>	0,79	<E→E→E→D→D→A→A>	0,79
<B→E→B→E→E→E→A>	0,95	<B→A→E→B→E→E→E>	0,84	<B→E→B→E→E→D→E>	0,79	<E→E→C→E→D→A→A>	0,79
<C→E→B→E→E→E→A>	0,95	<C→A→E→B→E→E→E>	0,84	<D→E→B→E→E→D→E>	0,79	<B→E→E→A→E→A→A>	0,79
<E→E→B→E→E→E→A>	0,95	<B→C→E→B→E→E→E>	0,84	<E→A→C→E→E→D→E>	0,79	<E→E→E→A→E→A→A>	0,79
<E→E→C→E→E→E→A>	0,95	<C→C→E→B→E→E→E>	0,84	<A→E→C→E→E→D→E>	0,79	<E→C→E→C→E→A→A>	0,79
<E→E→D→E→E→E→A>	0,95	<D→E→A→C→E→E→E>	0,84	<D→E→C→E→E→D→E>	0,79	<C→E→E→C→E→A→A>	0,79
<A→A→E→E→E→E→A>	0,95	<E→C→B→C→E→E→E>	0,84	<C→E→D→E→E→D→E>	0,79	<B→E→D→E→A→A>	0,79
<D→A→E→E→E→E→A>	0,95	<E→A→C→C→E→E→E>	0,84	<A→B→E→E→E→D→E>	0,79	<E→A→B→E→E→A→A>	0,79
<D→B→E→E→E→E→A>	0,95	<B→E→D→C→E→E→E>	0,84	<D→B→E→E→E→D→E>	0,79	<E→B→B→E→E→A→A>	0,79
<E→B→E→E→E→E→A>	0,95	<A→A→E→C→E→E→E>	0,84	<B→C→E→E→E→D→E>	0,79	<B→E→B→E→E→A→A>	0,79

Tabela 6.2: Sequências frequentes máximas com dados semanais

Sequência	Suporte	Sequência	Suporte
<D→E→A→E>	0,89	<A→E→D→E→A>	0,79
<D→D→D→A→A→D>	0,84	<D→A→A→E→E→A>	0,79
<D→A→E→A→D>	0,84	<A→D→A→A→D→A>	0,79
<E→D→A→A→E→A>	0,84	<D→E→A→A→D→A>	0,79
<A→D→E→E→A>	0,84	<E→E→A→D→A>	0,79
<D→A→A→B→A→A>	0,84	<D→A→A→D→D→A>	0,79
<A→D→D→D→A→A>	0,84	<E→A→D→D→A>	0,79
<A→A→D→E→A→A>	0,84	<A→E→D→D→A>	0,79
<E→A→A→A→A→A→E>	0,79	<E→A→E→D→A>	0,79
<A→D→A→A→A→E>	0,79	<A→A→E→E→D→A>	0,79
<A→A→E→A→E>	0,79	<D→A→E→E→D→A>	0,79
<A→D→A→E→D>	0,79	<E→A→A→B→A>	0,79
<D→A→E→D→D>	0,79	<E→D→A→B→A>	0,79
<A→B→D>	0,79	<A→A→A→A→A→A>	0,79
<A→D→A→A→A→D>	0,79	<D→A→A→A→A→A>	0,79
<D→D→A→A→A→D>	0,79	<E→A→A→A→A→A>	0,79
<D→A→A→D→A→D>	0,79	<E→D→A→A→A→A>	0,79
<A→E→D→A→D>	0,79	<D→E→A→A→A→A>	0,79
<E→E→D→A→D>	0,79	<A→D→D→A→A→A>	0,79
<A→A→E→A→D>	0,79	<A→E→D→A→A→A>	0,79
<A→C>	0,79	<E→E→D→A→A→A>	0,79
<A→E→B>	0,79	<E→A→D→A→A>	0,79
<A→D→A→A→B>	0,79	<A→D→A→B→A>	0,79
<D→D→A→A→B>	0,79	<D→A→A→A→D→A>	0,79
<E→D→A→A→B>	0,79	<D→D→A→A→D→A>	0,79
<D→E→A→A→B>	0,79	<A→D→A→D→A>	0,79
<A→A→A→A→A→E→A>	0,79	<E→D→A→D→A>	0,79
<D→A→A→A→A→E→A>	0,79	<A→A→D→D→A>	0,79
<E→A→A→A→A→E→A>	0,79	<A→E→E→D→A>	0,79
<A→D→A→A→E→A>	0,79	<A→A→A→A→E→A>	0,79
<D→D→D→A→E→A>	0,79	<D→A→A→A→E→A>	0,79
<A→E→A→E→A>	0,79	<D→D→A→A→E→A>	0,79
<A→B→E→A>	0,79	<A→D→A→E→A>	0,79
<E→A→D→E→A>	0,79	<E→D→A→E→A>	0,79

Tabela 6.3: Sequências frequentes máximas da Adidas, Allianz e Commerzbank

Adidas		Allianz		Basf		Commerzbank	
Sequências	Suporte	Sequências	Suporte	Sequências	Suporte	Sequências	Suporte
D→B→E→E	0,98	E→A→E→E	0,98	D→D→D→E	0,98	A→A→E→E	0,98
D→D→E→E	0,98	A→C→E→E	0,98	A→C→A→E	0,98	A→A→D→E	0,98
C→D→D→E	0,98	A→A→C→E	0,98	B→D→A→E	0,98	E→A→B→E	0,98
C→D→B→E	0,98	A→A→A→E	0,98	D→A→E→D	0,98	E→D→B→E	0,98
B→A→A→E	0,98	E→D→A→E	0,98	E→C→E→D	0,98	A→A→A→E	0,98
D→E→E→D	0,98	A→E→A→E	0,98	E→D→D→D	0,98	E→E→A→E	0,98
A→C→D→D	0,98	E→E→A→E	0,98	C→E→D→D	0,98	A→B→E→D	0,98
A→A→C→D	0,98	D→A→E→D	0,98	A→D→B→D	0,98	C→C→D→D	0,98
A→D→C→D	0,98	A→E→E→D	0,98	B→C→A→D	0,98	C→D→D→D	0,98
E→A→B→D	0,98	E→A→D→D	0,98	D→D→E→C	0,98	A→C→C→D	0,98
A→C→A→D	0,98	A→D→D→D	0,98	C→A→B→C	0,98	B→E→C→D	0,98
C→A→D→C	0,98	D→D→D→D	0,98	A→A→A→C	0,98	C→A→A→D	0,98
A→A→A→C	0,98	A→E→A→D	0,98	C→A→D→B	0,98	D→B→E→C	0,98
D→C→D→B	0,98	A→A→A→C	0,98	E→C→A→B	0,98	D→D→D→C	0,98
A→A→A→B	0,98	E→A→E→B	0,98	C→A→E→A	0,98	B→E→D→B	0,98
A→E→A→B	0,98	A→A→E→A	0,98	D→B→D→A	0,98	D→E→A→B	0,98
A→A→E→A	0,98	E→A→E→A	0,98	A→E→D→A	0,98	E→A→E→A	0,98
B→E→E→A	0,98	C→E→E→A	0,98	A→A→C→A	0,98	A→D→E→A	0,98
A→B→D→A	0,98	D→D→D→A	0,98	A→B→C→A	0,98	A→E→E→A	0,98
E→E→D→A	0,98	A→E→D→A	0,98	D→E→C→A	0,98	E→C→D→A	0,98
A→A→B→A	0,98	E→E→D→A	0,98	D→A→A→A	0,98	D→D→C→A	0,98
D→A→A→A	0,98	D→A→A→A	0,98	E→A→A→A	0,98	A→A→A→A	0,98
E→A→A→A	0,98	D→D→A→A	0,98	E→D→A→A	0,98	D→A→A→A	0,98
A→B→A→A	0,98	E→D→A→A	0,98	A→E→A→A	0,98	D→C→A→A	0,98
E→D→A→A					0,98	C→D→A→A	0,98
E→E→A→A							

Tabela 6.4: Sequências frequentes máximas da Continental, Daimler, Deutsche Bank e Deutsche Telekom

Continental		Daimler		Deutsche Bank		Deutsche Telekom	
Sequência	Suporte	Sequência	Suporte	Sequência	Suporte	Sequência	Suporte
A→A→E→E	0,98	A→A→A→E	0,98	A→E→D→E	0,98	A→E→B→E	0,98
D→A→A→E	0,98	B→D→A→E	0,98	A→E→B→E	0,98	A→A→A→E	0,98
A→B→A→E	0,98	D→A→E→D	0,98	A→B→A→E	0,98	E→D→A→E	0,98
E→D→A→E	0,98	A→D→E→D	0,98	D→D→A→E	0,98	B→E→A→E	0,98
B→A→E→D	0,98	E→D→D→D	0,98	D→A→E→D	0,98	A→A→E→D	0,98
A→D→D→D	0,98	D→E→D→D	0,98	D→C→D→D	0,98	D→A→E→D	0,98
D→D→D→D	0,98	A→A→B→D	0,98	A→D→C→D	0,98	A→C→D→D	0,98

D→A→B→D	0,98	B→A→B→D	0,98	D→A→B→D	0,98	C→D→D→D	0,98
A→A→A→D	0,98	A→A→A→D	0,98	A→B→A→D	0,98	C→A→C→D	0,98
A→E→A→D	0,98	D→A→A→B	0,98	A→E→A→D	0,98	A→C→A→C	0,98
E→E→A→D	0,98	A→B→A→B	0,98	E→A→D→C	0,98	A→E→A→C	0,98
E→A→D→B	0,98	D→D→A→B	0,98	A→B→D→C	0,98	A→A→E→B	0,98
A→D→A→B	0,98	A→A→E→A	0,98	B→A→E→B	0,98	E→A→E→A	0,98
D→D→A→B	0,98	A→A→D→A	0,98	C→A→A→B	0,98	E→B→E→A	0,98
D→A→E→A	0,98	A→B→D→A	0,98	E→A→A→B	0,98	D→D→D→A	0,98
A→E→E→A	0,98	D→D→D→A	0,98	A→D→A→B	0,98	A→E→D→A	0,98
A→A→D→A	0,98	A→E→D→A	0,98	D→E→A→B	0,98	E→A→C→A	0,98
E→A→D→A	0,98	D→A→B→A	0,98	E→B→E→A	0,98	A→A→A→A	0,98
A→B→D→A	0,98	A→A→A→A	0,98	E→D→E→A	0,98	D→A→A→A	0,98
D→D→D→A	0,98	D→A→A→A	0,98	B→A→D→A	0,98	D→D→A→A	0,98
A→E→D→A	0,98	E→A→A→A	0,98	C→D→D→A	0,98	E→D→A→A	0,98
D→A→B→A	0,98	B→D→A→A	0,98	B→D→C→A	0,98		
D→A→A→A	0,98	E→D→A→A	0,98	A→A→B→A	0,98		
A→D→A→A	0,98	A→E→A→A	0,98	E→A→B→A	0,98		
B→D→A→A	0,98			D→C→A→A	0,98		
				B→E→A→A	0,98		

Tabela 6.5: Sequências frequentes máximas da Fresenius, Heidelberg, Linde e Merck

Fresenius		Heidelberg		Linde		Merck	
Sequência	Suporte	Sequência	Suporte	Sequência	Suporte	Sequência	Suporte
A→A→E→E	0,98	A→A→E→E	0,98	A→D→E→E	0,98	E→A→E→E	0,98
A→E→E→E	0,98	C→A→E→E	0,98	C→A→D→E	0,98	A→A→B→E	0,98
E→E→C→E	0,98	D→D→D→E	0,98	E→E→D→E	0,98	C→A→A→E	0,98
C→A→A→E	0,98	D→A→A→E	0,98	A→A→C→E	0,98	E→D→A→E	0,98
C→D→A→E	0,98	A→B→A→E	0,98	E→B→B→E	0,98	B→E→A→E	0,98
D→A→E→D	0,98	A→C→A→E	0,98	B→B→E→D	0,98	A→B→E→D	0,98
A→A→C→D	0,98	E→D→A→E	0,98	B→E→E→D	0,98	A→E→E→D	0,98
A→A→A→D	0,98	D→D→E→D	0,98	D→A→B→D	0,98	A→A→B→D	0,98
A→B→A→D	0,98	A→E→E→D	0,98	C→C→A→D	0,98	D→A→E→C	0,98
E→C→E→C	0,98	A→D→D→D	0,98	E→D→E→C	0,98	E→E→D→C	0,98
E→E→E→C	0,98	D→A→B→D	0,98	D→E→C→C	0,98	A→A→A→C	0,98
A→A→D→C	0,98	E→A→A→D	0,98	B→A→A→C	0,98	A→B→D→B	0,98
E→A→A→C	0,98	E→A→A→C	0,98	D→E→E→B	0,98	A→A→C→B	0,98
B→A→D→B	0,98	B→D→A→B	0,98	A→A→B→B	0,98	C→A→A→B	0,98
A→D→C→B	0,98	E→D→A→B	0,98	A→B→B→B	0,98	E→A→A→B	0,98
C→B→A→B	0,98	B→A→E→A	0,98	B→B→B→B	0,98	A→A→E→A	0,98
A→E→E→A	0,98	D→A→E→A	0,98	E→E→B→B	0,98	B→E→D→A	0,98

A→C→D→A	0,98	A→A→D→A	0,98	A→A→A→B	0,98	E→D→C→A	0,98
A→E→D→A	0,98	A→B→D→A	0,98	E→D→A→B	0,98	A→E→C→A	0,98
C→E→C→A	0,98	D→E→D→A	0,98	A→B→D→A	0,98	B→D→B→A	0,98
B→A→B→A	0,98	E→E→D→A	0,98	B→E→D→A	0,98	A→A→A→A	0,98
D→C→B→A	0,98	A→A→C→A	0,98	E→C→C→A	0,98	B→A→A→A	0,98
D→A→A→A	0,98	D→A→B→A	0,98	B→B→B→A	0,98	D→B→A→A	0,98
E→C→A→A	0,98	A→D→A→A	0,98	D→A→A→A	0,98	D→C→A→A	0,98
E→D→A→A	0,98	A→E→A→A	0,98	B→B→A→A	0,98	E→C→A→A	0,98
E→E→A→A	0,98			B→D→A→A	0,98	A→E→A→A	0,98

Tabela 6.6: Sequências frequentes máximas da Siemens, Volkswagen, S&P500 e USA Oil Fund

Siemens		Volkswagen		S&P500		USA Oil Fund	
Sequência	Supporte	Sequência	Supporte	Sequência	Supporte	Sequência	Supporte
A→A→E→E	0,98	E→A→B→E	0,98	B→A→E→E	0,98	D→C→D→E	0,98
B→A→E→E	0,98	A→A→A→E	0,98	E→B→E→E	0,98	D→D→D→E	0,98
E→A→E→E	0,98	E→A→A→E	0,98	E→D→E→E	0,98	E→D→B→E	0,98
E→A→D→E	0,98	D→D→A→E	0,98	E→E→E→E	0,98	A→D→A→E	0,98
A→A→A→E	0,98	E→D→A→E	0,98	A→E→D→E	0,98	D→A→E→D	0,98
C→A→A→E	0,98	A→A→E→D	0,98	E→E→C→E	0,98	D→B→E→D	0,98
E→B→A→E	0,98	D→A→E→D	0,98	A→A→B→E	0,98	C→D→E→D	0,98
E→E→A→E	0,98	A→D→E→D	0,98	B→E→B→E	0,98	D→D→E→D	0,98
A→E→E→D	0,98	E→D→D→D	0,98	C→E→B→E	0,98	C→A→D→D	0,98
E→C→D→D	0,98	D→E→D→D	0,98	B→B→A→E	0,98	A→D→D→D	0,98
C→D→D→D	0,98	E→A→A→D	0,98	E→D→A→E	0,98	E→D→D→D	0,98
A→E→C→D	0,98	A→E→A→B	0,98	D→A→E→D	0,98	D→E→D→D	0,98
D→A→B→D	0,98	A→A→E→A	0,98	B→E→E→D	0,98	C→D→C→D	0,98
E→E→A→D	0,98	D→A→E→A	0,98	E→E→E→D	0,98	D→D→C→D	0,98
D→D→D→C	0,98	A→B→E→A	0,98	A→E→E→C	0,98	E→C→A→D	0,98
A→A→E→B	0,98	A→A→D→A	0,98	A→B→E→B	0,98	D→D→A→D	0,98
E→D→A→B	0,98	D→D→D→A	0,98	E→C→E→B	0,98	D→C→D→C	0,98
A→E→E→A	0,98	A→E→D→A	0,98	D→E→E→B	0,98	D→D→D→C	0,98
A→B→D→A	0,98	A→A→A→A	0,98	E→E→D→B	0,98	A→E→D→B	0,98
E→E→D→A	0,98	D→A→A→A	0,98	E→E→B→B	0,98	D→A→D→A	0,98
D→D→C→A	0,98	E→D→A→A	0,98	B→A→A→B	0,98	D→D→D→A	0,98
A→E→B→A	0,98	A→E→A→A	0,98	E→E→D→A	0,98	B→E→D→A	0,98
D→A→A→A	0,98	B→E→A→A	0,98	E→B→B→A	0,98	E→E→C→A	0,98
D→C→A→A	0,98			E→D→B→A	0,98	E→D→A→A	0,98
B→D→A→A	0,98			D→B→A→A	0,98		

Tabela 6.7: Sequências frequentes máximas da BMW, da Bayer e da Lufthansa

BMW		Bayer		Lufthansa	
Sequência	Suporte	Sequência	Suporte	Sequência	Suporte
C→D→D→E	0,98	A→A→D→E	0,98	A→B→D→E	0,98
A→A→C→E	0,98	E→A→D→E	0,98	D→D→B→E	0,98
A→A→A→E	0,98	D→E→D→E	0,98	B→B→A→E	0,98
E→A→A→E	0,98	C→A→A→E	0,98	D→B→D→D	0,98
A→A→E→D	0,98	E→A→A→E	0,98	A→D→D→D	0,98
D→C→D→D	0,98	A→D→A→E	0,98	D→D→D→D	0,98
D→D→C→D	0,98	E→D→A→E	0,98	B→A→B→D	0,98
D→A→B→D	0,98	A→A→E→D	0,98	A→D→B→D	0,98
A→E→A→D	0,98	D→A→E→D	0,98	D→A→A→D	0,98
A→D→D→C	0,98	A→D→E→D	0,98	A→E→A→D	0,98
A→A→A→C	0,98	A→A→A→D	0,98	B→E→A→D	0,98
E→A→A→C	0,98	A→E→A→D	0,98	B→D→E→B	0,98
E→D→A→B	0,98	A→A→E→C	0,98	E→A→D→B	0,98
A→A→E→A	0,98	D→A→E→A	0,98	B→D→D→B	0,98
A→C→E→A	0,98	E→D→E→A	0,98	D→A→B→B	0,98
D→D→E→A	0,98	A→A→D→A	0,98	D→E→B→B	0,98
E→A→D→A	0,98	A→E→D→A	0,98	B→B→A→B	0,98
A→B→D→A	0,98	A→E→C→A	0,98	A→D→A→B	0,98
A→E→D→A	0,98	D→A→A→A	0,98	B→A→E→A	0,98
A→A→C→A	0,98	E→A→A→A	0,98	D→B→E→A	0,98
C→A→A→A	0,98	E→C→A→A	0,98	A→A→D→A	0,98
D→A→A→A	0,98	E→D→A→A	0,98	D→D→D→A	0,98
A→C→A→A	0,98	A→E→A→A	0,98	A→B→B→A	0,98
B→D→A→A	0,98	D→E→A→A	0,98	E→B→B→A	0,98
C→E→A→A	0,98			D→D→A→A	0,98
D→E→A→A	0,98				

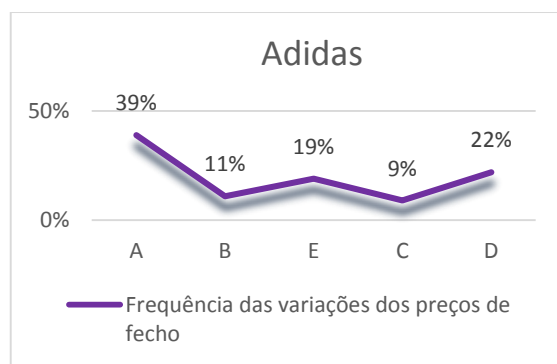


Figura 6.1: Frequências das variações dos preços de fecho na Adidas

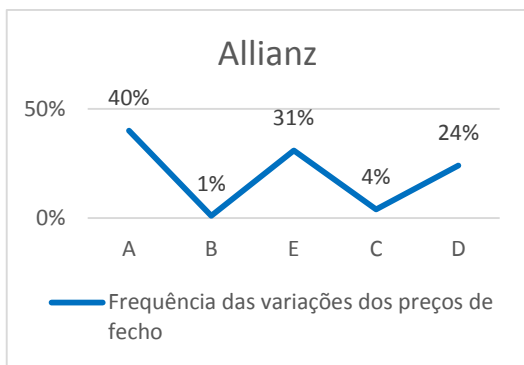


Figura 6.2: Frequências das variações dos preços de fecho na Allianz

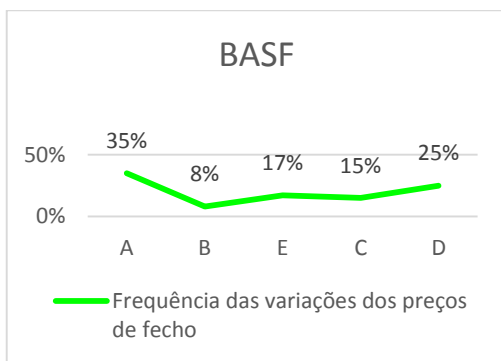


Figura 6.3: Frequências das variações dos preços de fecho na BASF

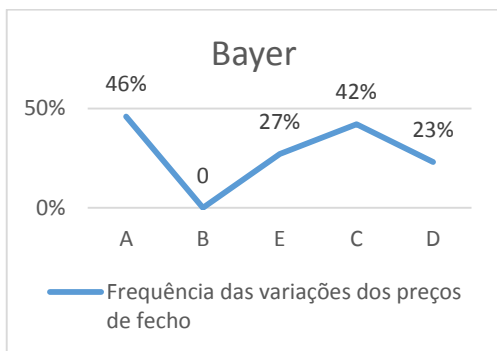


Figura 6.4: Frequências das variações dos preços de fecho na Bayer

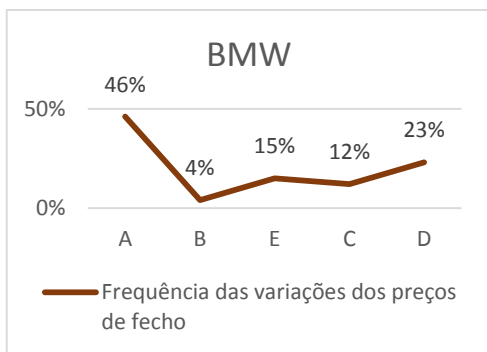


Figura 6.5: Frequências das variações dos preços de fecho na BMW

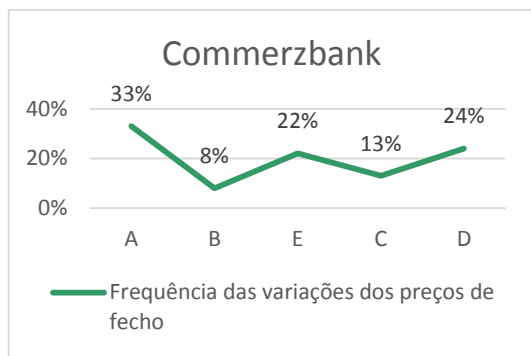


Figura 6.6: Frequências das variações dos preços de fecho no Commerzbank

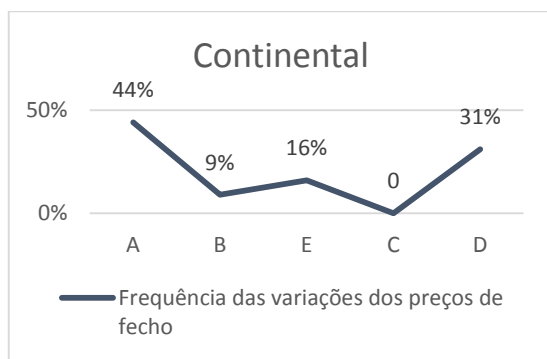


Figura 6.7: Frequências das variações dos preços de fecho no Continental

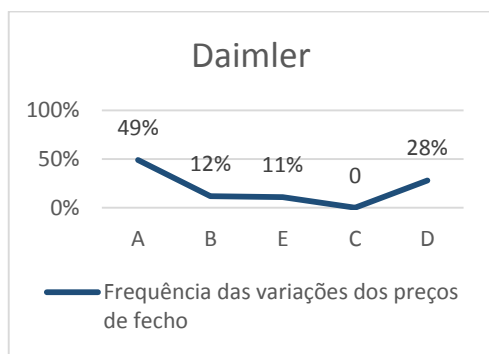


Figura 6.8: Frequências das variações dos preços de fecho na Daimler

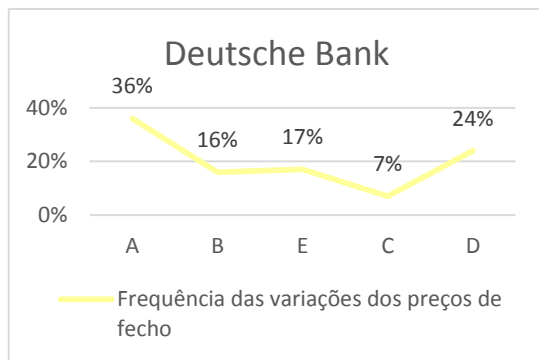


Figura 6.9: Frequências das variações dos preços de fecho no Deutsche Bank

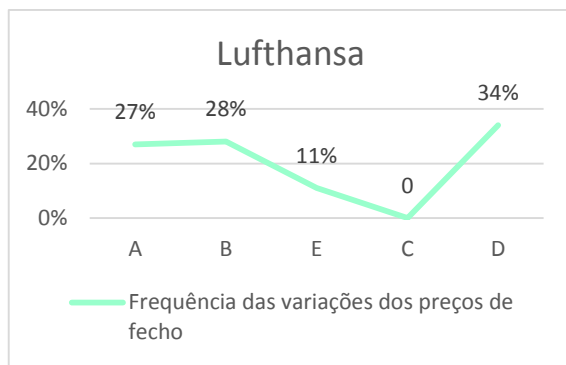


Figura 6.10: Frequências das variações dos preços de fecho na Lufthansa

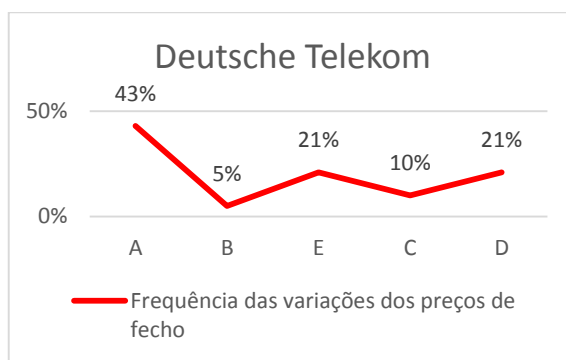


Figura 6.11: Frequências das variações dos preços de fecho na Deutsche Telekom

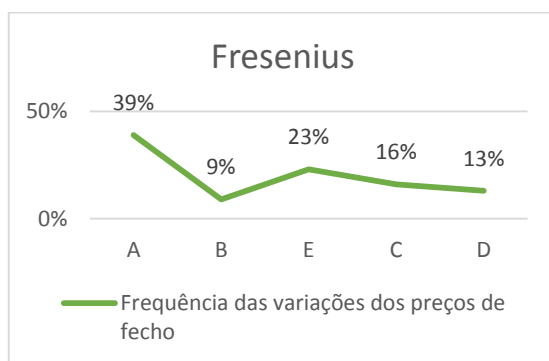


Figura 6.12: Frequências das variações dos preços de fecho na Fresenius

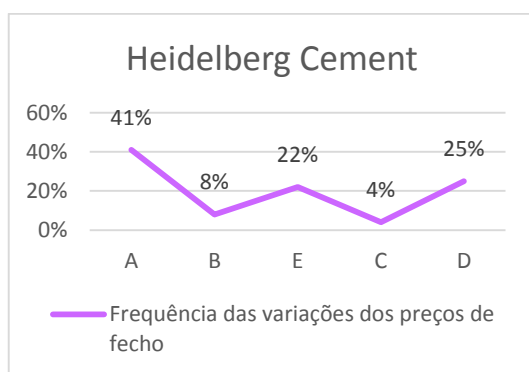


Figura 6.13: Frequências das variações dos preços de fecho na Heidelberg Cement

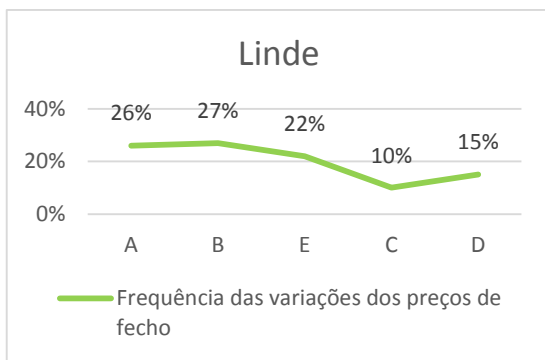


Figura 6.14: Frequências das variações dos preços de fecho na Linde

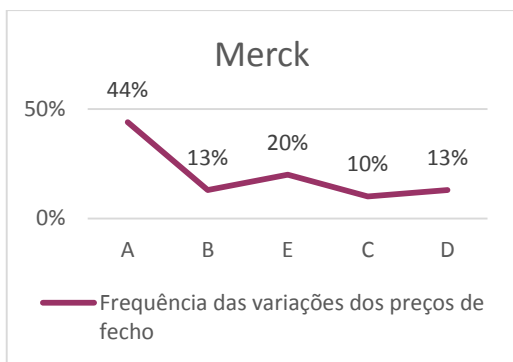


Figura 6.15: Frequências das variações dos preços de fecho na Merck

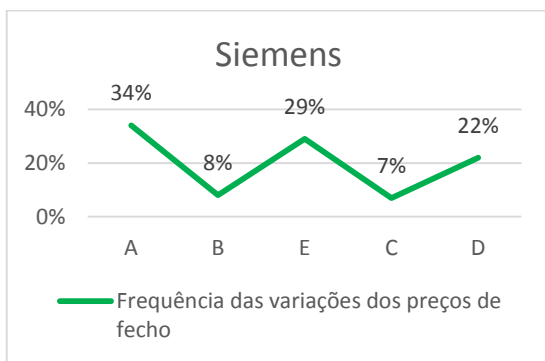


Figura 6.16: Frequências das variações dos preços de fecho na Siemens

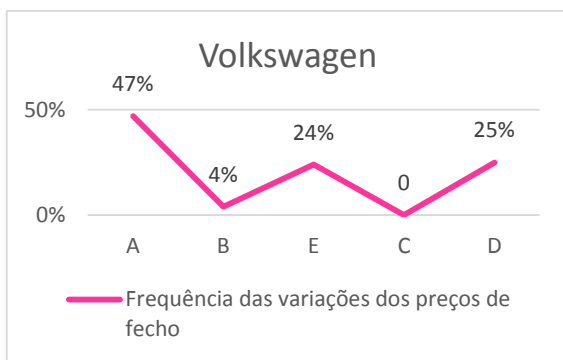


Figura 6.17: Frequências das variações dos preços de fecho na Volkswagen

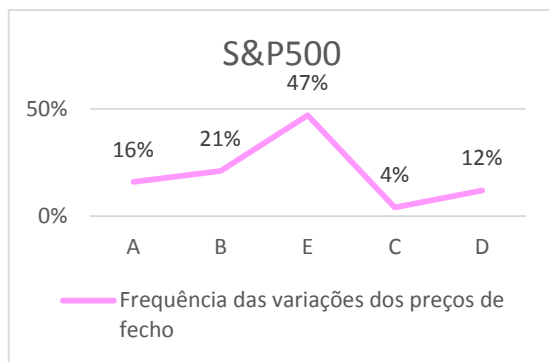


Figura 6.18: Frequências das variações dos preços de fecho na S&P500

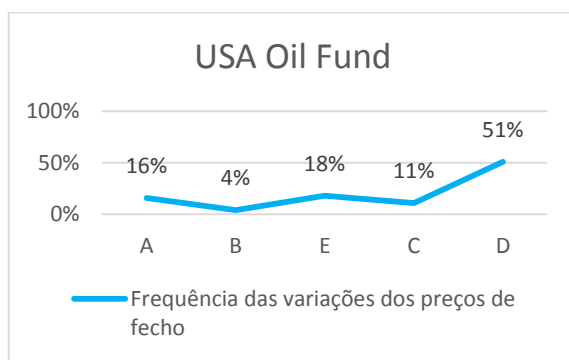


Figura 6.19: Frequências das variações dos preços de fecho no USA Oil Fund